JOURNAL OF
Econometrics

ELSEVIER Journal of Econometrics 73 (1996) 5–59

Long memory processes and fractional integration in econometrics

Richard T. Baillie

Department of Economics, Michigan State University, East Lansing, MI 48824, USA

Abstract

This paper provides a survey and review of the major econometric work on long memory processes, fractional integration, and their applications in economics and finance. Some of the definitions of long memory are reviewed, together with previous work in other disciplines. Section 3 describes the population characteristics of various long memory processes in the mean, including ARFIMA. Section 4 is concerned with estimation and examines semiparametric procedures in both the frequency and time domain, and also the properties of various regression based and maximum likelihood techniques. Long memory volatility processes are discussed in Section 5, while Section 6 discusses applications in economics and finance. The paper also has a concluding section.

Key words: Fractional integration; Long memory processes; Hurst effect; ARFIMA processes; FIGARCH processes; Stochastic volatility
JEL classification: C22

1. Introduction

This article provides a review of the growing literature of econometric work on long memory, fractionally integrated processes that are associated with

I am greatly indebted to the editor, Max King, and to three anonymous referees for their very helpful comments on an earlier version of this paper. I also wish to acknowledge several useful discussions with Clive Granger and particularly helpful comments from Tim Bollerslev, Ching-Fan Chung, David Dickey, Frank Diebold, Hans-Ole Mikkelsen, Franz Palm, Peter Robinson, John Rogers, Philip Rothman, Peter Schmidt, and Casper de Vries. This article has also benefited from comments made by seminar participants at Duke University, Federal Reserve Bank of St. Louis, University of California–San Diego, University of California–Santa Barbara, University of California–Riverside, University of Southern California, and University of Washington.

SSDI 0304-4076(95)01732-S

hyperbolically decaying autocorrelations and impulse response weights. The article attempts to be quite comprehensive in terms of coverage of results that appear of direct relevance for econometricians and also of the increasing number of applications in economics and finance. The review of the statistical literature in this field is deliberately selective given the econometric orientation of the survey.

While long memory models have only really been used by econometricians since around 1980, they have played a role in the physical sciences since at least 1950, with statisticians in fields as diverse as hydrology and climatology long recognizing the presence of long memory within data recorded over both time and space. The presence of long memory can be defined from an empirical, data-oriented approach in terms of the persistence of observed autocorrelations. The extent of the persistence is consistent with an essentially stationary process, but where the autocorrelations take far longer to decay than the exponential rate associated with the ARMA class. This phenomenon has been noted in different data sets by Hurst (1951, 1957), Mandelbrot and Wallis (1968), Mandelbrot (1972), and McLeod and Hipel (1978) among others. When viewed as the time series realization of a stochastic process, the autocorrelation function exhibits persistence that is neither consistent with an I(1) process nor an I(0) process. One of the most compelling motivations concerning the importance of long memory, fractionally integrated processes is related to the rate of decay associated with the impulse response coefficients of a process. The classical theory of stationary time series, and indeed many of the models used in econometrics, requires the existence of the Wold decomposition. The conditions for the Wold decomposition are relatively weak and, apart from the possible presence of a purely deterministic component, little more than square summability and martingale behavior is required for the innovation sequence associated with the stochastic component. A considerable amount of success in econometrics has been obtained from using the ARMA class of models which impose an exponential, or geometric, rate of decay on the Wold decomposition coefficients. This strategy has taken time series econometrics a long way theoretically and also in modeling empirical behavior. However, there is no conceptual reason for restricting attention to exponential rates of decay in the Wold decomposition, and there are indeed both theoretical and economic reasons for considering slower rates, such as hyperbolic decay. While a considerable amount of recent work has emphasized the role of persistence of shocks, most of it has been directed towards testing for the presence of unit roots in autoregressive representations of univariate and vector processes. However, the knife-edge distinction between I(0) and I(1) processes can be far too restrictive. The fractionally differenced process can be regarded as a halfway house between the I(0) and I(1) paradigms. One attraction of long memory models is that they imply different long run predictions and effects of shocks to conventional macroeconomic approaches.

There is considerable evidence on the success of applying long memory models to time series data in the physical sciences, and rather less to macroeconomics where in many cases it seems hard to distinguish I(d) behavior from I(1) behavior. However, there is substantial evidence that long memory processes describe rather well financial data such as forward premiums, interest rate differentials, and inflation rates. Perhaps the most dramatic empirical success of long memory processes has been in recent work on modeling the volatility of asset prices and power transformations of returns. In this context the approach has yielded hitherto unknown empirical regularities, which have spawned possible new insights into understanding market behavior and the pricing of risk.

Apart from 38 background references, this paper cites 138 further articles concerned with long memory processes. Of these, 20 are in the field of probability theory and 82 are statistical/econometric papers concerning population characteristics and/or inference. There are 36 application papers: 10 in macroeconomics, 18 in finance, and 8 in the geophysical sciences.

2. Preliminary ideas and definitions

2.1. Data considerations and the Hurst effect

The origin of interest in long memory processes appears to have come from the examination of data in the physical sciences and preceded interest from economists. Perhaps the most well-known example has been in hydrology, and has included tidal flows and the inflows into reservoirs and was originally documented by Hurst (1951). The articles by Hurst (1951, 1956) analyze 900 geophysical time series and were partly motivated through the desire to understand the persistence of streamflow data and the design of reservoirs. A survey of various stochastic models used in the analysis of river flows is in Lawrance and Kottegoda (1977); while Hipel and McLeod (1978a, b) and McLeod and Hipel (1978) consider various climatological applications. Long memory models may also be of interest for investigating the possibility of climatic change. Several authors have noted an apparent upward trend in world temperatures since the second half of the nineteenth century (e.g., Seater, 1993). An important policy issue concerns whether the higher temperatures are evidence of climatic change and global warming brought on by the man made emissions of green house gasses, or whether the recent observed temperatures are merely part of the regular cyclical variation that is known to occur in world temperature readings. Hurst (1951) and Mandelbrot and Wallis (1968) were also aware that temperature data and tree ring series exhibited long memory characteristics.

Fig. 1 presents a graph of annual tree ring measurements from Mount Campito and extends from 3436 BC through 1969 AD, a total of 5405

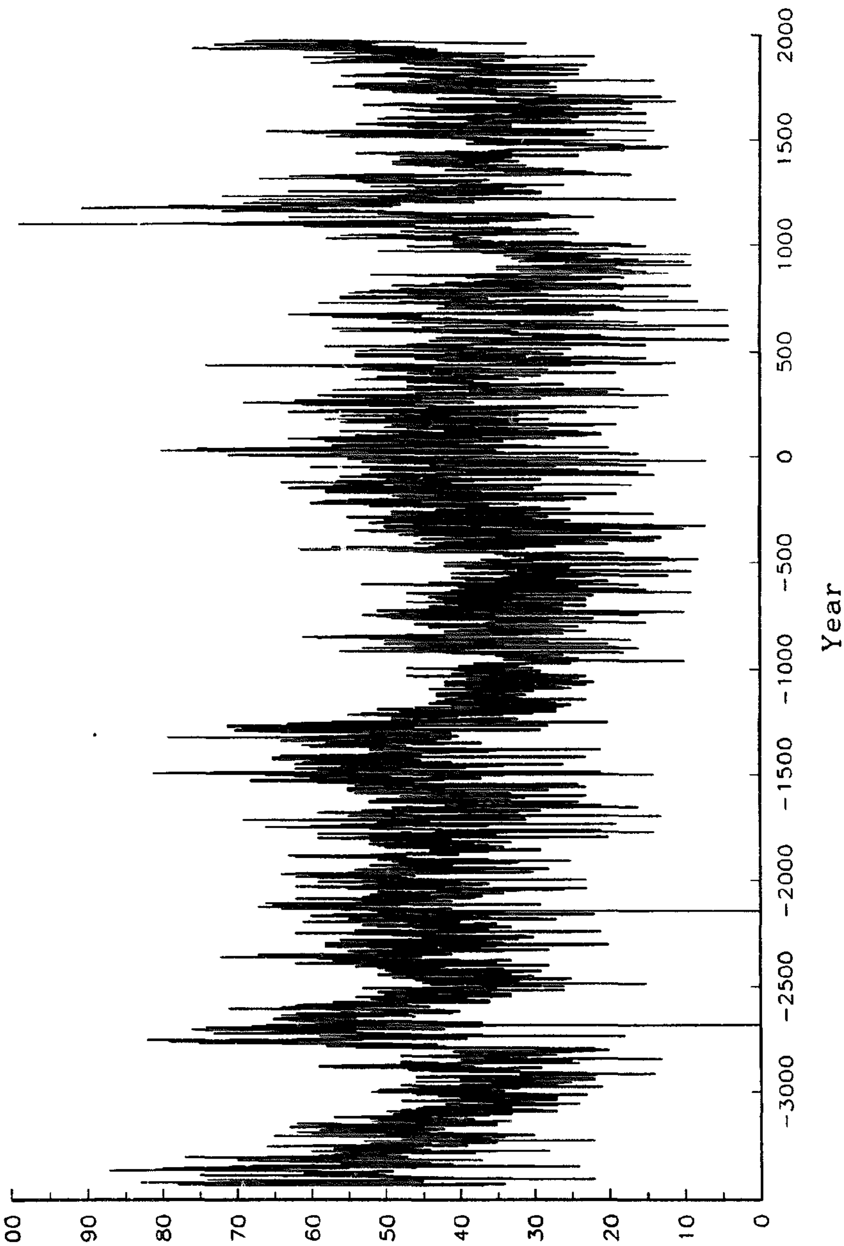

Fig. 1. Mount Campito tree ring data.

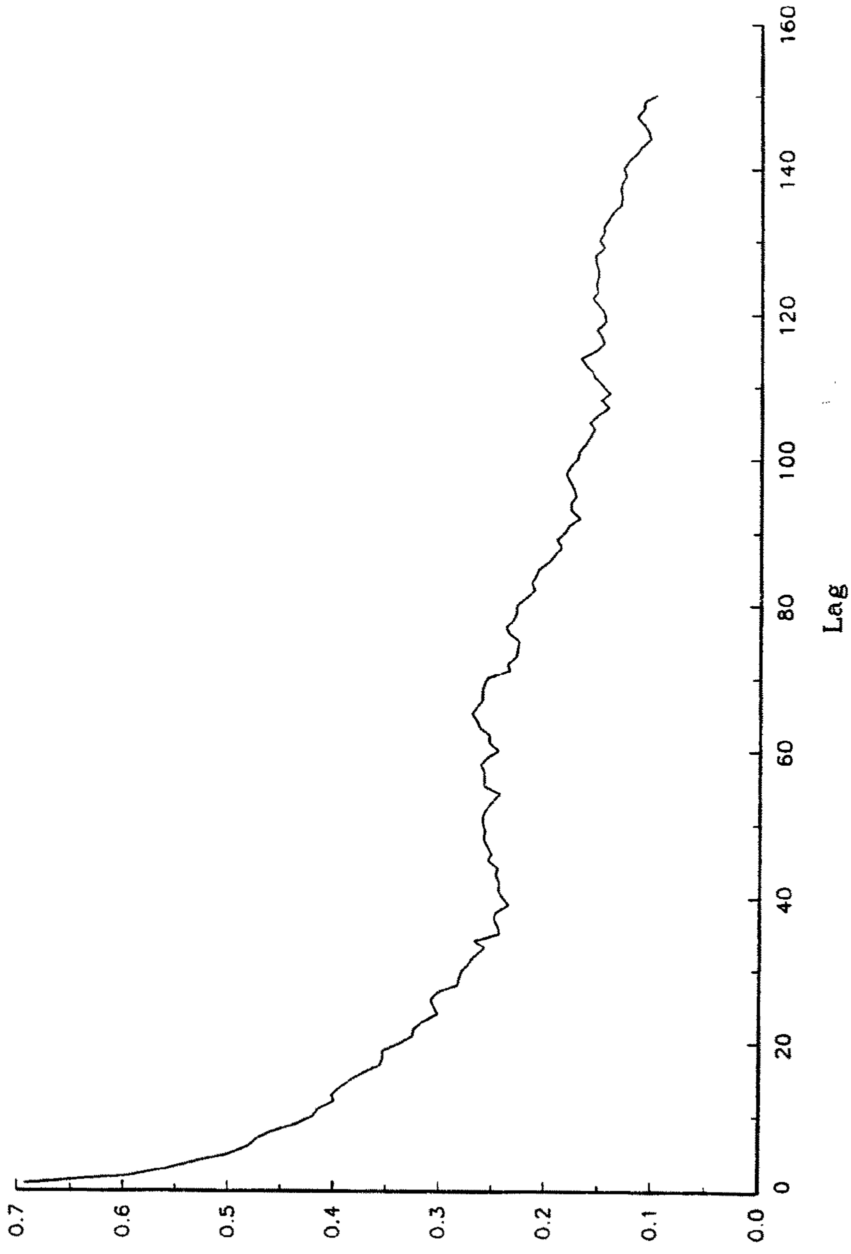

Fig. 2. Autocorrelations of tree rings.

observations.[1] This series is relatively typical of much of the geophysical data that first motivated Hurst (1951, 1956) in his formulation of the concept of long memory. The correlogram of this series in Fig. 2 shows that the autocorrelations exhibit a clear pattern of slow decay and persistence. All the first 150 years autocorrelations are positive and considerably greater than 0.027, their asymptotic standard error.

Fig. 3 graphs the well-known Beveridge (1925) Wheat Price Index which gives annual price data from 1500 through 1869, averaged over many locations in western and central Europe. The autocorrelations are graphed in Fig. 4. One of the characteristics of long memory series is that the autocorrelations of the original series frequently have the appearance of being nonstationary, while the differenced series can appear over differenced. This property seems true of the Beveridge wheat price series and also of the US monthly Consumer Price Index (CPI) inflation series. The autocorrelations of these series for both levels and differences are given in Table 1.

Authors such as Whittle (1956) and Beran (1989, 1992a) discuss the occurrence of long memory within a spatial context. For example, Whittle (1956) considers persistent autocorrelation discovered in plots of 'independently' treated land, while Beran (1992a) discusses measurements of the 1-kg standard weight by the US National Bureau of Standards in Washington, DC, where despite the close to ideal conditions correlations between measurements appear to decay according to a hyperbolic law.

2.2. *Definitions of long memory*

There are several possible definitions of the property of 'long memory'. Given a discrete time series process y_t with autocorrelation function ρ_j at lag j, then according to McLeod and Hipel (1978), the process possesses long memory if the quantity

$$\lim_{n\to\infty} \sum_{j=-n}^{n} |\rho_j| \tag{1}$$

is nonfinite. Equivalently, the spectral density $f(\omega)$ will be unbounded at low frequencies. A stationary and invertible ARMA process has autocorrelations which are geometrically bounded, i.e., $|\rho_k| \leqslant cm^{-k}$, for large k, where $0 < m < 1$ and is hence a short memory process. Fractionally integrated processes, to be discussed in Section 3 onwards are long memory processes given the definition in (1). In particular, the process y_t is said to be integrated of order d, or $\mathrm{I}(d)$, if

$$(1 - L)^d y_t = u_t, \tag{2}$$

[1] I am grateful to Ian McLeod of the University of Western Ontario for providing this data.

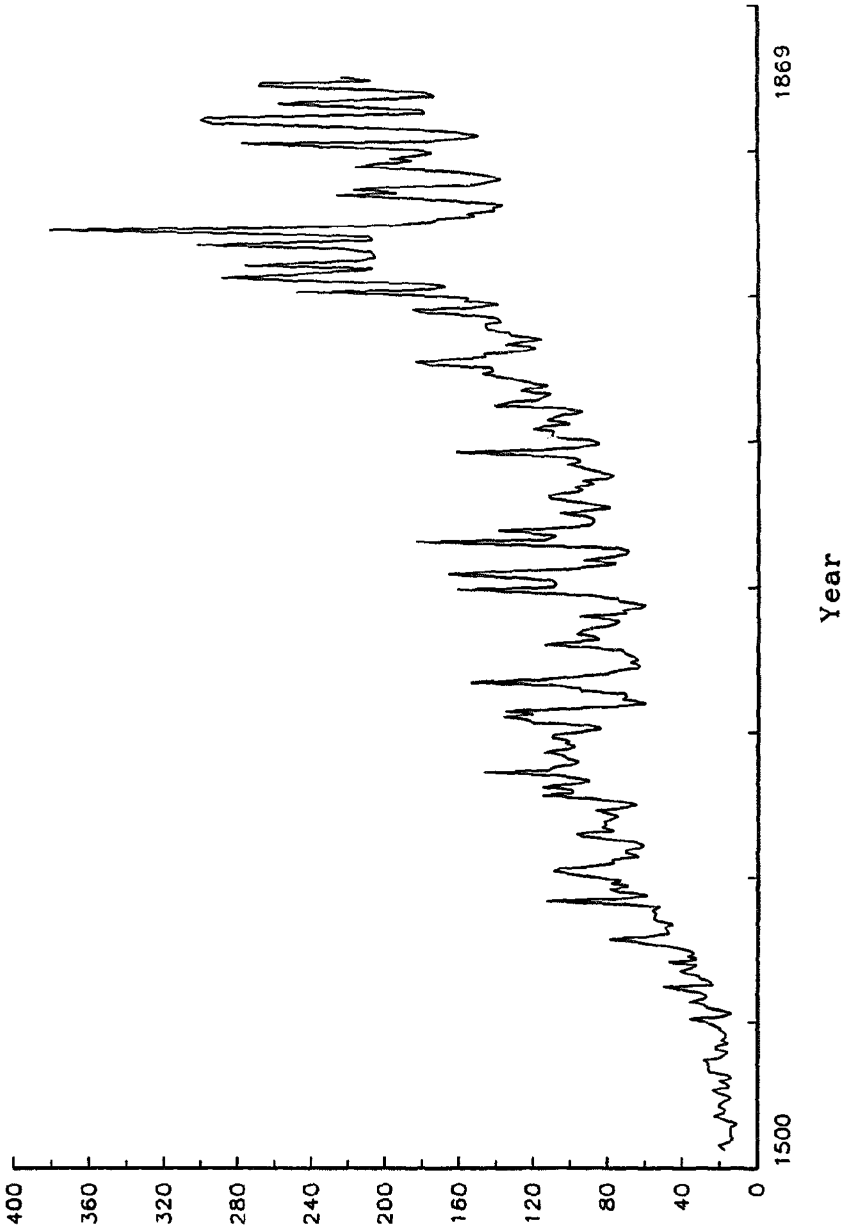

Fig. 3. Beveridge wheat price index.

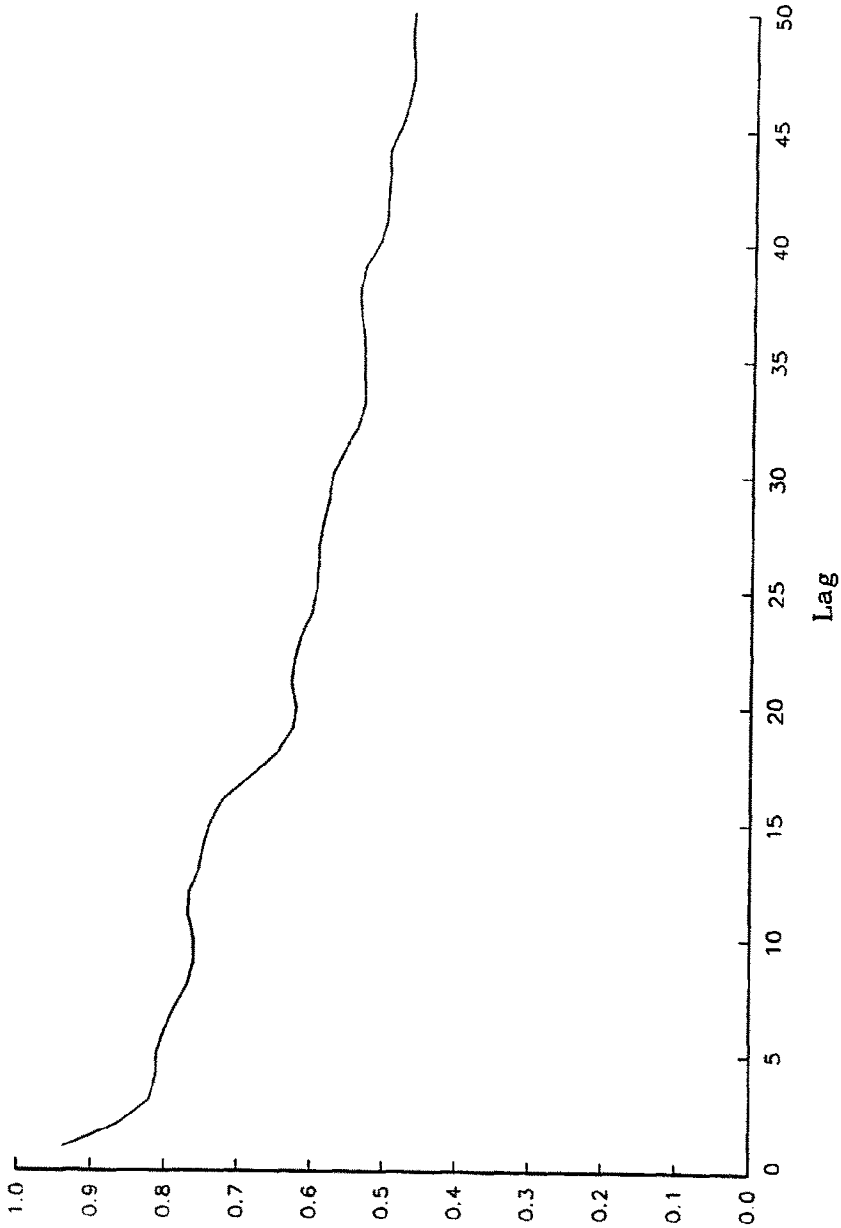

Fig. 4. Autocorrelations of Beveridge wheat price index.

Table 1
Autocorrelations of the US Consumer Price Index and the Beveridge Wheat Price Index, Consumer Price Index data from January 1948 through July 1990, $T = 523$

US Consumer Price Index									
1	0.467	2	0.423	3	0.399	4	0.360	5	0.316
6	0.305	7	0.312	8	0.359	9	0.386	10	0.349
11	0.320	12	0.278	13	0.234	14	0.180	15	0.211
16	0.229	17	0 160	18	0.129				
Differenced US Consumer Price Index									
1	−0.499	2	0.112	3	−0.086	4	0.032	5	−0.050
6	−0.013	7	−0.024	8	0.028	9	0.042	10	−0.029
11	0.024	12	0.037	13	−0.034	14	−0.089	15	0.073
16	0.008	17	0.084	18	−0.036				
Beveridge Wheat Price Index									
1	0.9367	2	0.8624	3	0.8200	4	0.8095	5	0.8097
6	0.7984	7	0.7850	8	0.7674	9	0.7587	10	0.7598
11	0.7663	12	0.7646	13	0.7514	14	0.7454	15	0.7368
16	0.7190	17	0.6812	18	0.6447	19	0.6243	20	0.6204
21	0.6264	22	0.6231	23	0.6144	24	0.6002	25	0.5925
26	0.5906	27	0.5901	28	0.5838	29	0.5771	30	0.5719
31	0.5568	32	0.5383	33	0.5294	34	0.5293	35	0.5299
36	0.5309	37	0.5359	38	0.5375	39	0.5280	40	0.5100
41	0.5018	42	0.4998	43	0.4984	44	0.4989	45	0.4854
46	0.4744	47	0.4678	48	0.4673	49	0.4696	50	0.4672
Differenced Beveridge Wheat Price Index									
1	0.1073	2	−0.3173	3	−0.2850	4	−0.0609	5	0.1212
6	0.0203	7	0.0200	8	−0.1001	9	−0.1024	10	−0.0196
11	0.0947	12	0.1036	13	−0.0988	14	−0.0181	15	0.0789
16	0.1802	17	0.0301	18	−0.1185	19	−0.1315	20	−0.0799
21	0.0766	22	0.0316	23	−0.0130	24	−0.0264	25	−0.0258
26	−0.0070	27	0.0509	28	−0.0059	29	−0.0074	30	0.0764
31	0.0223	32	−0.0688	33	−0.0617	34	−0.0013	35	0.0039
36	−0.0399	37	0.0219	38	0.0699	39	0.0622	40	−0.0732
41	−0.0739	42	0.0175	43	0.0019	44	0.1398	45	−0.0223
46	−0.0459	47	−0.0666	48	−0.0201	49	0.0334	50	0.0721

where L is the lag operator, $-0.5 < d < 0.5$, and u_t is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies. One important class of process occurs when u_t is I(0) and is covariance stationary. For $0 < d < 0.5$, the process is long memory in the sense of the condition (1), its autocorrelations are all positive and decay at a hyperbolic rate. For $-0.5 < d < 0$, the sum of absolute values of the processes autocorrelations tends to a constant, so that it has short memory according to definition (1). In this situation the ARFIMA$(0, d, 0)$ process is said to be 'antipersistent' or to have 'intermediate memory', and all its autocorrelations, excluding lag zero, are negative and decay hyperbolically to zero.

Alternatively, the memory of a process y_t can be expressed in terms of the behavior of its partial sum

$$S_T = \sum_{t=1}^{T} y_t. \tag{3}$$

Rosenblatt (1956) defines short range dependency in terms of a process that satisfies strong mixing, so that the maximal dependence between two points of a process becomes trivially small as the distance between these points increases. More concretely, a process y_t can be defined as having short memory if

$$\sigma^2 = \lim_{T \to \infty} \mathrm{E}(T^{-1} S_T^2) \tag{4}$$

exists and is nonzero, and

$$[1/\sigma T^{1/2}] S_{[rT]} \Rightarrow B(r) \quad \text{for all} \quad r \in [0, 1], \tag{5}$$

where $[rT]$ is the integer part of rT, $B(r)$ is standard Brownian motion, and $\Rightarrow$ denotes convergence in distribution. This allows departures from covariance stationarity, but requires the existence of moments up to a certain order.

A wider definition of long memory is to include any process which possesses an autocovariance function for large k, such that

$$\gamma_k \approx \Xi(k) k^{2H-2}, \tag{6}$$

where $\approx$ denotes approximate equality for large k and where $\Xi(k)$ is any slowly varying function at infinity[2] and is described in detail by Resnick (1987). Helson and Sarason (1967) show that any process with $H > 0$ and autocovariance function given by (6) violates the strong mixing condition, and hence is long memory or long range dependent.

Taqqu (1975) studies the weak convergence of a linear combination of a 'long memory' type process, where the weights are functions of Hermite polynomials. Specifically, the results are for a stochastic process, $\sum_{t=1,[Np]} H_m(y_t)$, where y_t is Gaussian with zero mean and an autocovariance function obeying (6), $0 \leqslant p \leqslant 1$, and H_m is the mth Hermite polynomial. For the case $H < [1 - (1/2m)]$ then an appropriately normalized version of $\sum_{t=1,[Np]} H_m(y_t)$,

[2] A function $f(x)$ is defined as being regularly varying at infinity with index α if $\lim_{t \to \infty} [f(tx)/f(t)] = x^{\alpha}$, for all $x > 0$, so that asymptotically $f(x)$ is a power function. The function is slowly varying at infinity if $\alpha = 0$, so that $f(x)$ asymptotically becomes a constant. $f(x) = \log(x)$ is an example of a slowly varying function at infinity.

will converge to Brownian motion. However, when $[1 - (1/2m)] < H < 1$, the limit depends on m, is non-Gaussian for $m \geqslant 2$, and when $m = 2$, the limit is the Rosenblatt process. Fox and Taqqu (1985) have provided further results for the quadratic form,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i-j} H_m(y_t) H_n(y_t),$$

where a_i are finite constants, $H_m(\cdot)$ again denotes the mth Hermite polynomial, and y_t is the same long memory process as before. Similarly, Fox and Taqqu find the normalized sum of the quadratic form converges either to Brownian motion or to a Rosenblatt process. Fox and Taqqu (1987), Giraitis and Surgailis (1990), and Beran and Terrin (1994) show that a vector of quadratic forms with long memory converges to a vector of independent Gaussian random variables. The constants in the quadratic form have to decay at sufficient speed to offset the long range dependencies in y_t.

2.3. Fractional Brownian motion

Regular Brownian motion is a continuous time stochastic process, $B(r)$, composed of independent Gaussian increments. Mandelbrot and Van Ness (1968) also note that in a sense fractional Brownian motion, $B_H(r)$, can be regarded as the approximate $(1/2 - H)$ fractional derivative of regular Brownian motion,

$$B_H(r) = [1/\Gamma(H + 1/2)] \int_0^r (r - x)^{H - 1/2} \mathrm{d}B(x) \quad \text{for} \quad r \in (0, 1), \tag{7}$$

where $\Gamma(\cdot)$ is the gamma function, $B(x)$ is regular Brownian motion with unit variance, and H is the Hurst coefficient, originally due to Hurst (1951). When $H = 1/2$, $B_H(r)$ reduces to regular Brownian motion, $B(r)$. The autocovariance function of fractional Brownian motion is given by

$$\mathrm{E}\,|B_H(t) - B_H(s)|^2 = |t - s|^{2H},$$

and

$$\gamma_k \approx |k|^{2H-2}, \tag{8}$$

so that for high lags hyperbolic decay occurs in the autocovariance function. Continuous time fractional noise is denoted by $B_H(t)'$ and is the derivative of fractional Brownian motion. The $(1/2 - H)$ fractional derivative of continuous time white noise reduces to white noise when $H = 1/2$.

Avram and Taqqu (1987) and Davydov (1970) develop a functional central limit theorem for fractionally integrated processes. In particular, when (2) is fractional white noise,

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

where $d < 0.5$ and ε_t is independent and identically distributed with zero mean and finite variance; and on defining S_T as

$$S_T = \sum_{t=1}^{[Tr]} y_t,$$

it can be shown that

$$(1/\sigma_T) S_{[Tr]} \Rightarrow B_d(r). \tag{9}$$

The above fractional Brownian motion $B_d(r)$ is defined, analogously to Eq. (7), as

$$B_d(r) = \{1/\Gamma(d+1)\} \int_{-0}^{r} (r - x)^d [\mathrm{d}B(x)], \tag{10}$$

where the fractional differencing parameter, d, is related to the Hurst coefficient as $d = H - 1/2$. One additional property is that the partial sum S_T in (3) of $B_H(r)$ variates is $O_p(T^H)$, while for a short memory process, S_T in (3) is $O_p(T^{1/2})$. Furthermore,

$$\sigma_T^2 = \mathrm{var}(S_T) \approx t^{2H} \Xi(T),$$

where $\Xi(T)$ is again a slowly varying function, such as log T.

A related literature on self-similar processes was developed by Kolmogorov (1940) in continuous time. The definition is considerably stronger than when defining particular processes with long range dependencies. Formally, a process y_t is self-similar with respect to a parameter H, if for any m and time points $t_1, \dots, t_m$, the joint distribution of $\{y_{t1}, \dots, y_{tm}\}$ is identical to α^{-H} times the joint distribution of $\{y_{\alpha t1}, \dots, y_{\alpha tm}\}$. Kolmogorov (1940) and Mandelbrot and Van Ness (1968) show that the autocovariance function of a self-similar process y_t, which is observed at discrete, regular intervals of time, is given by

$$\gamma_k = C(1/2)\gamma_0\{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\}, \tag{11}$$

for $C > 0$ and $1/2 < H < 1$. For high lags k,

$$\gamma_k \approx CH(2H - 1)|k|^{2(H-1)}, \tag{12}$$

as $k \to \infty$. Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1969a) refer to the process generating the above autocovariance function as fractional Gaussian noise, and Sinai (1976) provides a result for the spectral density of the process. In some studies the degree of persistence is considered to be related to the time interval of observation. In Mandelbrot's (1972, 1975) terminology, a 'fractal' is defined as $|\Delta y_t| = (\Delta t/c)^H$, where Δt is the time interval and c is a constant. A self-similar fractal has the same H, or Hurst coefficient, for all choices of time intervals.

Another self-similar process, which is otherwise unrelated to long memory processes concerns random variables with densities exhibiting excess kurtosis. Fat-tailed densities are self-similar with respect to their tail behavior and extreme value theory (as discussed for example by Leadbetter, Lindgren, and Rootzen, 1983) is concerned with the limiting distribution of the order statistics. For example any convolution of a Student t density with degree of freedom α will have the same tail index and is said to be self-similar with respect to tail behavior. Hols and de Vries (1991) and Koedijk, Schafgans, and de Vries (1990) find this property to be apparent in exchange rate returns.[3]

3. Theoretical models

3.1. Fractional white noise

While the discrete time analog of Brownian motion is the random walk, the discrete time version of fractional Brownian motion is fractionally differenced white noise. The process was independently developed by Granger (1980), Granger and Joyeux (1980), and Hosking (1981), although earlier work by Adenstedt (1974) and Taqqu (1975) shows an awareness of the representation. The process is defined as

$$(1 - L)^d (y_t - \mu) = \varepsilon_t, \tag{13}$$

where $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t^2) = \sigma^2$, and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for $s \neq t$, and where the fractional parameter d is possibly noninteger. It will be seen that the process is weakly stationary for $d < (1/2)$ and is invertible for $d > -(1/2)$. The infinite-order autoregressive representation of fractional white noise is given by

$$y_t = \sum_{k=0}^{\infty} \pi_k y_{t-k} + \varepsilon_t, \tag{14}$$

where the infinite-order autoregressive representation weights π_k are obtained from the binomial expansion,

$$(1 - L)^d = \{1 - dL + d(d-1)L^2/2! - d(d-1)(d-2)L^3/3! + \cdots\},$$

[3] One interesting related study by de Haan (1990) uses limit laws for the distribution of a maximum from a sample, to study the distribution of high tides and the appropriate dyke levels in the Netherlands. Extremal values were associated with the devastating high tides of 1570 and that of 1953, which caused over 2000 deaths in the Netherlands. de Haan first attempts to remove temporal dependencies from the data before using extremal analysis.

for any real $d > -1$. The expansion can also be represented in terms of the hypergeometric function,

$$(1 - L)^d = \sum_{k=0}^{\infty} \Gamma(k - d)L^k/\Gamma(k + 1)\Gamma(-d) = F(-d, 1, 1; L), \tag{15}$$

for $d > 0$, and where $F(a, b; c; z)$ is the hypergeometric function defined as

$$F(a,b;c;z) = \Gamma(c)/[\Gamma(a)\Gamma(b)] \times \sum_{i=1}^{\infty} z^i\Gamma(a + i)\Gamma(b + i)/[\Gamma(c + i)\Gamma(i + 1)]. \tag{16}$$

The typical autoregressive coefficient at lag k, given by π_k, is

$$\pi_k = \{d(d - 1)(d - 2) \ldots (d - k + 1)(-1)^k\}/k! \tag{17}$$
$$= \{(-d)(1 - d)(2 - d) \ldots (k - 1 - d)\}/k!,$$

and since

$$\Gamma(k - d) = \{(k - d - 1)(k - d - 2) \ldots (2 - d)(1 - d)(-d)\}\Gamma(-d),$$

it follows that the infinite autoregressive representation coefficients can be expressed as

$$\pi_k = \Gamma(k - d)/\{\Gamma(-d)\Gamma(k + 1)\}. \tag{18}$$

Similarly, the fractional white noise process can be expressed as an infinite-order moving average representation, or Wold decomposition,

$$y_t = \sum_{k=0}^{\infty} \psi_k\varepsilon_{t-k} \tag{19}$$
$$= (1 - L)^{-d}\varepsilon_t$$
$$= \{1 + dL + d(d + 1)L^2/2! + d(d + 1)(d + 2)L^3/3! + \ldots\}\varepsilon_t. \tag{20}$$

Since

$$\Gamma(d + k) = d(d + 1)(d + 2) \ldots (d + k - 1)/\Gamma(d),$$

it follows that

$$\psi_k = \Gamma(k + d)/\{\Gamma(d)\Gamma(k + 1)\}. \tag{21}$$

The cumulative impulse response is the total impact of a unit innovation and is given by $\psi(1) = \sum_{j=0,\infty} \psi_j$, and the spectral density at the zero frequency is $f(0) = \psi(1)^2\sigma^2$ for $d < 0$ and $f(0) = \infty$ for $d > 0$.

Brockwell and Davis (1987) show that y_t is convergent in mean square through its spectral representation. Also, since $\sum_{j=0,\infty} \psi_j^2 < \infty$, the fractional white noise process is mean square summable and stationary for $d < 0.5$. When $d = 0.5$, the ARFIMA(0,0.5,0) process is a discrete time version of '$1/f$' noise and is just nonstationary, since $\psi_k \approx k^{-1/2}$, and hence $\sum \psi_k^2$ just fails to converge.

Odaki (1993) discusses invertibility in the sense that the MSE of the one-step-ahead linear predictor from a finite-order AR(p) converges to the innovation variance. 'Invertibility' in this sense was originally discussed by Granger and Andersen (1978), and Odaki (1993) shows that $d > -1$ is a sufficient condition for the ARFIMA process.

The most important properties of the fractional white noise process are summarized in Table 2 and were all derived by Granger (1980), Granger and Joyeux (1980), and Hosking (1981). Of particular interest are the long-run

Table 2
Properties of fractional white noise

Infinite MA representation coefficients	*Asymptotic approximation*
$\psi_k = \Gamma(k+d)/\{\Gamma(d)\Gamma(k+1)\}$	
$\psi_k = \psi_{k-1}\{(k-1+d)/k\}$	$\psi_k \approx \{\Gamma(d)\}^{-1}k^{d-1}$
$\psi_k = \prod_{0 \leq j \leq k}[(j-1+d)/j]$	
Infinite AR representation coefficients	
$\pi_k = \Gamma(k-d)/\{\Gamma(-d)\Gamma(k+1)\}$	
$\pi_k = \pi_{k-1}\{(k-1-d)/k\}$	$\pi_k \approx \{\Gamma(-d)\}^{-1}k^{-d-1}$
$\pi_k = \prod_{0 \leq j \leq k}[(j-1-d)/j]$	
Autocovariances	
$\gamma_0 = \sigma^2\Gamma(1-2d)/\{\Gamma^2(1-d)\}$	
$\gamma_k = \{\sigma^2\Gamma(k+d)\Gamma(1-2d)\}/\{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)\}$	
$\gamma_k = \{(\sigma^2/2\pi)\sin(\pi d)[\Gamma(k+d)\Gamma(1-2d)]\}/\{\Gamma(k+1-d)\}$	
$\gamma_k = (-1)^k\Gamma(1-2d)/[\Gamma(1-k-d)\Gamma(k+1-d)]$	
Autocorrelations	
$\rho_1 = d/(1-d)$, $\rho_2 = d(1+d)/\{(1-d)(2-d)\}, \ldots$	
$\rho_k = \prod_{0<i\leq k}\{(i-1+d)/(i-d)\}$	
$\rho_k = \{\Gamma(k+d)\Gamma(1-d)\}/\{\Gamma(k-d+1)\Gamma(d)\}$	$\rho_k = \{\Gamma(1-d)/\Gamma(d)\}k^{2d-1}$
Partial autocorrelations	
$\phi_{kk} = d/(k-d)$ for $k = 1,2$	
Power spectrum	
$f(\omega) = (\sigma^2/2\pi)[1-e^{-i\omega}|^{-2d}$ and at low frequencies $f(\omega) \approx c\omega^{-2d}$	
$f(\omega) = (\sigma^2/2\pi)[2\sin(\omega/2)]^{-2d}$	$f(0) < \infty$ if $d \leqslant 0$

properties of some of the characteristics of the process. On using Stirling's approximation for large k that $\Gamma(k+a)/\Gamma(k+b) \approx k^{a-b}$, it can be established that $\psi_k \approx c_1 k^{d-1}$, $\pi_k \approx c_2 k^{-d-1}$, and $\rho_k \approx c_3 k^{2d-1}$, where the c_i are constants. Hence the impulse response weights, infinite autoregressive coefficients, and autocorrelation coefficients all exhibit slow hyperbolic decay for large k.

Also, the power spectrum $f(\omega)$ of the process is closely approximated at low frequencies, as $\omega \to 0$, by

$$f(\omega) \approx \omega^{-2d},$$

compared with $f(\omega) \approx \omega^{-2}$ for a I(1), unit root process. Hence fractional white noise is consistent with the 'typical spectral shape' of many economic time series originally noted by Granger (1966), and the ARFIMA model to be discussed in the next section can be useful in representing spectral density functions at low frequencies. This is in contrast to differenced series which have a power spectrum that is close to being zero at low frequencies. Also, for any constant c and low frequency ω, $f(\omega) = |c|^{2d} f(c\omega)$, so that the process is self-similar at low frequencies with the properties of y_t remaining invariant to the time interval.

A further property discussed by Sowell (1990) concerns the behavior of the contiguous sum S_T in (3), when y_t is fractional white noise as in (13). Then Sowell (1990) shows that

$$\operatorname{var}(S_T) = \sigma^2 \Gamma(1-2d)\{(1+2d)\Gamma(1+d)\Gamma(1-d)\}^{-1} c, \tag{22}$$

where

$$c = [\Gamma(1+d+T)/\Gamma(T-d) - \Gamma(1+d)/\Gamma(-d)]$$

and

$$\lim_{T\to\infty} [\operatorname{var}(S_T)]\, T^{-(1+2d)} = \sigma^2 \Gamma(1-2d)\{(1+2d)\Gamma(1+d)\Gamma(1-d)\}^{-1}.$$

Hence,

$$\operatorname{var}(S_T) = \mathrm{O}(T^{2d+1}), \tag{23}$$

which implies that the variance of the partial sum of an I(0) process, with $d = 0$, grows linearly, i.e., at a rate of $\mathrm{O}(T)$. For a process with intermediate memory with $-0.5 < d < 0$, the variance of the partial sum grows at a slower rate than the linear rate, while for a long memory process with $0 < d < 0.5$, the rate of growth is faster than a linear rate. Diebold (1989) considers a possible test for the presence of I(d) behavior based on the variance time function $R(k)$,

$$R(k) = k\sigma_1^2/\sigma_k^2, \tag{24}$$

for positive integer valued k, and where $\sigma_k^2 = \operatorname{var}(y_t - y_{t-k})$ and $\sigma_k^2 \sim \mathrm{O}(k^{2d-1})$ for an I(d) process. If $d < 1/2$, the variance time function becomes flat; for

$1/2 < d < 1$, then $R(k)$ grows at a decreasing rate, and for $1 < d < 3/2$, then $R(k)$ will grow at an increasing rate. Diebold (1989) tabulates the fractiles of $R(k)$.

3.2. The ARFIMA process

An important and more flexible class of process in discrete time has been introduced by Granger and Joyeux (1980), Granger (1980, 1981), and Hosking (1981), and is the ARFIMA(p,d,q) model

$$\phi(L)(1-L)^d(y_t-\mu)=\theta(L)\varepsilon_t, \tag{25}$$

where d denotes the fractional differencing parameter, all the roots of $\phi(L)$ and $\theta(L)$ lie outside the unit circle, and ε_t is white noise. The y_t process defined by Eq. (25) and for $d \neq 0$ is then said to be I(d). The Wold decomposition and autocorrelation coefficients will all exhibit a very slow rate of hyperbolic decay. For $-0.5 < d < 0.5$, the process is covariance stationary, while $d < 1$ implies mean reversion. For an I(d) process, the spectral density is such that $f(0) = 0$ for $d < 0$ and $f(0) = \infty$ for $d > 0$. For small frequencies, ω, an approximation for $d > 0$ is given by $f(\omega) \approx \omega^{-2d}$, while the process has infinite variance for $d > 0.5$. In particular, for the fractional process, Sowell (1986, 1992a) shows that

$$\gamma_k=\sigma^2\sum_{j=1}^{p}\xi_j\sum_{n=0}^{q}\sum_{m=0}^{q}\theta_n\theta_m C(d,d,p+n-m-k,\lambda_j), \tag{26}$$

where λ_j is the jth root of the AR polynomial lag operator, $\phi(L)$, and

$$\xi_j=\left[\lambda_j\prod_{i=1,p}(1-\rho_i\rho_j)\prod_{k=1,p;k\neq i}(\rho_i-\rho_k)\right]^{-1},$$

$$C(w,v,k,\rho)=G(w,v,k)[\rho^{2p}F(v+k,1;1-w+k;\rho)$$
$$+F(w-k,1;1-v-k;\rho)-1],$$

$$G(w,v,k)=[\Gamma(1-w-v)\Gamma(v+k)]/[\Gamma(1-w+k)\Gamma(1-v)\Gamma(v)].$$

Parametric expressions for the infinite autoregressive and infinite moving average representation weights for the general ARFIMA(p,d,q) process are relatively complicated functions of the hypergeometric function. Chung (1994a) provides some alternative methods for the calculation of these autocorrelations. In particular,

$$\gamma_k=\sum_{j=-q}^{q}a_j\sum_{n=1}^{p}\theta_n C(k-p-j;\psi_n),$$

where

$$C(k-p-j;\psi_k)=\psi_k^{2p}\sum_{m=0}^{\infty}\psi_k^m\gamma^*_{k-p-j-m}+\sum_{n=1}^{\infty}\psi_k^n\gamma^*_{k-p-j+n},$$

γ_k^* is the autocovariance at lag k of an ARFIMA$(0, d, 0)$ process,

$$a_k = \left[\psi_k \prod_{i=1,p} (1 - \psi_i \psi_k) \prod_{m \neq k,p} (\psi_k - \psi_m)\right]^{-1}, \qquad k = 1, p,$$

$$\psi_j = \sum_{i=0}^{q-|j|} \theta_i \theta_{i+|j|}.$$

For high lags, hyperbolic decay is also evident in the autocorrelations of the ARFIMA process and

$$\rho_k \approx c k^{2d-1}, \tag{27}$$

where $c > 0$. Also, the spectral density function is

$$\begin{aligned} f(\omega) &= (\sigma^2/2\pi)|\theta(e^{-i\omega})|^2 |\phi(e^{-i\omega})|^{-2} |1 - e^{-i\omega}|^{-2d} \\ &= (\sigma^2/2\pi)|\theta(e^{-i\omega})|^2 |\phi(e^{-i\omega})|^{-2} [2|1 - \cos(\omega)|]^{-2d}, \end{aligned} \tag{28}$$

and for low frequencies as $\omega \to 0$,

$$f(\omega) \approx (\sigma^2/2\pi)[\theta(1)/\phi(1)]^2 \omega^{-2d}. \tag{29}$$

It is often important to derive the impulse response weights from the ARFIMA(p, d, q) process in (25). Following Campbell and Mankiw (1987), the impulse response weights are defined by first differencing y_t in (25), to obtain

$$(1 - L) y_t = A(L) \varepsilon_t,$$

where

$$A(L) = (1 - L)^{1-d} \phi(L)^{-1} \theta(L). \tag{30}$$

The lag polynomial $A(L)$ can be expressed in terms of the hypergeometric function as

$$A(L) = F(d - 1, 1, 1; L) \phi(L)^{-1} \theta(L),$$

and from Gradszteyn and Ryzhnik (1980, pp. 1039–1042), $F(d - 1, 1, 1; L) = 0$ for $d < 1$. Hence, for $d < 1$,

$$A(1) = F(d - 1, 1, 1; 1) \phi(1)^{-1} \theta(1) = 0. \tag{31}$$

The impact of a unit innovation at time t on the process y_{t+k} is then given as $1 + \sum_{j=1,k} A_j$. For a mean reverting process, $A(1) = 0$. For any process $y_t \sim I(d)$ and for $d < 1$, it follows from (31) that y_t will be mean reverting. While y_t will not be covariance stationary for $0.5 < d < 1$, it will nevertheless still be mean reverting.

3.3. Aggregation and ARFIMA

Apart from the inherent reasonableness of having impulse response weights exhibiting slow hyperbolic decay, an alternative explanation has been provided by Robinson (1978) and Granger (1980). Suppose individual agents have stable AR(1) processes with coefficient $\alpha_j(\lambda)$, where λ indexes the population and $\alpha(\lambda)$ is a random variable with distribution function G and is independent of all innovation processes. Granger (1980) uses the beta distribution in the context of contemporaneous aggregation of panel data and obtained the power law of behavior of the corresponding unconditional autocovariances with rate k^{2d-1}. Mandelbrot (1971) suggests a similar idea in the context of Monte Carlo simulation of fractionally integrated time series. Granger (1980) considers

$$z_t = \sum_{i=1}^{N} y_{it},$$

which is the aggregate of N component and independent processes, y_{it}, such that for $i = 1, \ldots, N$,

$$y_{it} = \alpha_i y_{it-1} + \varepsilon_{it}. \tag{32}$$

For small N, z_t is ARMA($N, N-1$). Granger (1980) then considers the autoregressive coefficients α_i, to be drawn from a beta (0, 1) distribution,

$$\mathrm{d}F(\alpha) = [2/B(p,q)]\alpha^{2p-1}(1-\alpha^2)^{q-1}, \quad 0 \leqslant \alpha \leqslant 1, \quad p > 0, \quad q > 0,$$

and shows that in the limit for large N,

$$z_t \sim \mathrm{I}(1 - q/2),$$

which implies fractional behavior for the aggregate process. If $q > 1$, then $(1 - q/2) > 1/2$, and z_t will have an infinite variance. Lin (1991) provides some further results on aggregation in the context of long memory processes.

3.4. *Prediction from ARFIMA processes*

Granger and Joyeux (1980) and Geweke and Porter-Hudak (1983) consider prediction from an ARFIMA process in (25) by using the infinite autoregressive representation

$$y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + \varepsilon_t, \tag{33}$$

where

$$\pi(L) = (1-L)^d \phi(L)\theta(L)^{-1}.$$

They consider prediction based on a version of (33) truncated after k lags. Peiris (1987) and Peiris and Perera (1988) discuss some of the formulae for calculating predictions from the autoregressive representation. Since the ARFIMA process is not compatible with any finite-dimensional state space representation, there is no readily available solution to the truncation problem associated with using the autoregressive representation for prediction. A further currently unresolved issue concerns the effect of parameter estimation in ARFIMA processes, and the extent to which this increases prediction uncertainty. These effects may well be substantial in small samples, and are an area worthy of further investigation. Ray (1993a) considers a related issue concerning the asymptotic prediction MSE from approximating fractional white noise with a finite-order AR(p) model with estimated parameters. She finds the quality of the approximation to be very sensitive to both the order of the approximating autoregression and to the forecast horizon.

3.5. Gegenbauer and generalized ARFIMA processes

Gray, Zhang, and Woodward (1989) consider the theoretical properties of Gegenbauer and related processes. The simplest case is the pure Gegenbauer process given by

$$(1 - 2\xi L + L^2)^{\lambda} y_t = \varepsilon_t. \tag{34}$$

The process is covariance stationary if (i) $|\xi| < 1$ and $0 < \lambda < 0.5$ or (ii) $|\xi| = 1$ and $\lambda < 0.25$. The process is invertible if (i) $|\xi| < 1$ and $\lambda > -0.50$ and (ii) $|\xi| = 1$ and $\lambda > -0.25$. When $\xi = 1$, the process in (34) reduces to fractional white noise with $d = 2\xi$. A stationary Gegenbauer process exhibits long memory in the form of long memory harmonic behavior in its autocorrelation function. In particular, the form of the autocorrelations is dependent on the region of the parameter space. The most interesting region is for $|\xi| < 1$ and $0 < \lambda < 0.5$, then

$$\rho_k \approx C'[\cos(k\omega_0)k^{2\lambda-1}], \tag{35}$$

where $\omega_0 = \cos^{-1}(\lambda)$ is known as the G frequency and determines the harmonic frequency and C' is a constant that is independent of λ and ω_0. In this part of the parameter space the process has spectral density of

$$f(\omega) = \sigma^2\{4[\cos(\omega) - \xi]^2\}^{-\lambda}. \tag{36}$$

When $\xi = 1$ and $\lambda = d/2$, the Gegenbauer process in (34) reduces to the fractional white noise or ARFIMA$(0,d,0)$ process. Hence for $\xi = 1$ and $0 < \lambda < 0.25$,

$$\rho_k = [\Gamma(1-2\lambda)\Gamma(k+2\lambda)]/[\Gamma(2\lambda)\Gamma(k-2\lambda+1)],$$

while for $k \to \infty$,

$$\rho_k \approx k^{4\lambda-1}.$$

For $\xi = -1$ and $0 < \lambda < 0.25$,

$$\rho_k = (-1)^k[\Gamma(1-2\lambda)\Gamma(k+2\lambda)]/[\Gamma(2\lambda)\Gamma(k-2\lambda+1)],$$

and as $k \to \infty$,

$$\rho_k \approx (-1)^k k^{4\lambda - 1}.$$

Chung (1996) discusses other properties of Gegenbauer processes and considers approximate maximum likelihood estimation of these processes. The properties of the so-called Gegenbauer ARMA process, or GARMA process, are considered by Gray, Zhang, and Woodward (1989) and Chung (1994); although the first mention of this type of process appears to have been by Hosking (1981). The process is defined as

$$\phi(L)(1 - 2\xi L + L^2)^\lambda (y_t - \mu) = \theta(L)\varepsilon_t, \tag{37}$$

where the Gegenbauer process is appended with an ARMA component. The GARMA process is stationary if $|\xi| < 1$, $\lambda < 0.25$ (i.e., $d < 0.50$) and is invertible if $|\xi| < 1$ and $\lambda > -0.25$ (i.e., $d > -0.50$). On defining $v = \cos^{-1}(\xi)$, then the spectral density function is given by

$$f(\omega) = (\sigma^2/2\pi)|\theta(e^{-i\omega})|^2 |\phi(e^{-i\omega})|^{-2} [2|\cos(v) - \cos(\omega)|]^{-2d}. \tag{38}$$

As $\omega \to v$, then the spectral density function in (38) becomes $f(\omega) \to |\omega^2 - v^2|^{-2d}$, so that the spectral density function is unbounded as $\omega \to v$. The form of the autocovariance function and the impulse response weights are known for lag $k \to \infty$ to be

$$\gamma_k \approx \cos(kv) k^{2d-1} \tag{39}$$

and

$$\psi_k \approx \cos[kv + d(v - 0.5\pi)] k^{d-1}. \tag{40}$$

A further parametric long memory process has been suggested by Porter-Hudak (1990), who considers the seasonal fractionally differenced process

$$(1 - L^s)^d y_t = \varepsilon_t, \tag{41}$$

where s is the seasonal period, and analogously to Eq. (19), the process will have an infinite moving average representation given by

$$y_t = \psi(L)\varepsilon_t,$$

where $\psi(L) = (1 - L^s)^{-d}$ and

$$\psi_{sk} = \Gamma(sk - d)/[\Gamma(sk + 1)\Gamma(-d)] \quad \text{for} \quad k = 1, 2, 3, \ldots,$$

$$\psi_{sk} = \sigma^2 \cos(k\pi/s)\Gamma(1 - 2d)/[\Gamma(1 - d + sk)\Gamma(1 - d - k/s)],$$

and is zero for other lagged values. Asymptotically for large lags k,

$$\psi_k \approx ck^{d-1}. \tag{42}$$

The autocovariance function at lag k is given by

$$\gamma_k = \sigma^2 \cos(k\pi/s)\Gamma(1-2d)/[\Gamma(1-d+k/s)\Gamma(1-d-k/s)]. \tag{43}$$

The rates of decay of the autoregressive representation weights, the autocorrelations and the infinite moving average representation coefficients exactly coincide with those of the fractional white noise process in Table 2. A more general seasonal ARFIMA, or ARFISMA, model is given by

$$\phi(L)(1-L^s)^d y_t = \theta(L)\varepsilon_t. \tag{44}$$

Closed form expressions for the autocorrelation function of the above process are currently unavailable. However, the spectrum of the ARFISMA model in (44) is given by

$$f(\omega) = (\sigma^2/2\pi)|\theta(e^{-i\omega})|^2|\phi(e^{-i\omega})|^{-2}\{2[1-\cos(s\omega)]\}^{-2d}. \tag{45}$$

The spectrum is unbounded at frequencies $\omega_j = (2\pi j)/s$, for $j = 0, 1, 2, \ldots, (s/2)$, so that the model contains a persistent trend and $(s/2)$ persistent cyclical components. Hence the ARFISMA process shows a behavior at seasonal frequencies similar to that of the ARFIMA process at the zero frequency. Ray (1993b) presents an example of using seasonal ARFIMA models to predict monthly revenue data.

3.6. *Fractional cointegration*

From Granger (1981, 1983), two time series, $y_t \sim \mathrm{I}(d)$ and $x_t \sim \mathrm{I}(d)$, are said to be fractionally cointegrated of order (d, b) if $z_t = (y_t - \beta x_t) \sim \mathrm{I}(d-b)$, where $d > (1/2)$ and $d \geqslant b > 0$. In general, the order of integration of a linear combination of component processes, will be the maximum of the component processes. Granger has also provided an error correction formulation for fractionally cointegrated processes. If $y_t \sim \mathrm{I}(d)$ is a k-dimensional vector and z_t is a set of cointegrating vectors such that $z_t = \alpha' y_t \sim \mathrm{I}(d-b)$, then Granger has shown the appropriate error correction representation to be

$$H(\mathrm{L})(1-L)^d y_t = -\gamma[1-(1-L)^b](1-L)^{d-b} z_t + C(L)\varepsilon_t, \tag{46}$$

where $H(0) = I$ and $C(1) < \infty$. The possible fractional cointegration of I(1) asset prices may be a common phenomenon and will be discussed later. If economic fundamentals are only important in the long run, then mean reversion of returns will only occur over a very long horizon and the error correction term from (46) may be useful in reducing prediction MSE of the martingale model over long forecast horizons. Robinson (1992) also considers a form of cointegration in this context.

4. Estimation and testing

4.1. The rescaled range statistic

The original statistical measurement of long memory due to Hurst (1951) and used by Mandelbrot (1972, 1975) is the rescaled range or R/S statistic. The rescaled range statistic R_T/s_T is defined as

$$R_T = \max_{0 \leq j \leq T} \left\{ \sum_{j=1}^{T} (y_j - j\bar{y}) \right\} - \min_{0 \leq j \leq T} \left\{ \sum_{j=1}^{T} (y_j - j\bar{y}) \right\}, \tag{47}$$

where R is the range, s_T is the sample standard deviation, and $\bar{y}$ is the sample mean,

$$s_T = \{(1/T) \sum (y_t - \bar{y})^2\}^{1/2}. \tag{48}$$

Hurst (1951), Mandelbrot and Wallis (1968), Mandelbrot and Taqqu (1979), Taqqu (1975, 1977), and Lo (1991) showed that

$$\operatorname*{plim}_{T \to \infty} \{T^{-H}(R_T/s_T)\} = \text{constant}. \tag{49}$$

The idea of R/S analysis introduced by Hurst (1951) is to very informally write the above as

$$\log[\mathrm{E}(R_T/s_T)] \approx \text{constant} + H[\log(T)],$$

and the Hurst coefficient H is then estimated as $\log[R_T/s_T]/[\log(T)]$, or alternatively by taking the slope coefficient of a regression of $\log[R_T/s_T]$ on $\log(t)$, for different values of t. Since a short memory process would have a value H equal to 1/2, an estimated value of H that exceeds 1/2 is interpreted as evidence of long memory. Various alternative methods for estimating H from the above relationship are discussed by Mandelbrot and Wallis (1968, 1969b) and Davies and Harte (1987). Lo (1991) shows that $T^{-1/2}R_T/s_T$ is asymptotically distributed as the range of a standard Brownian Bridge on the unit interval and has an expectation of $(\pi/2)^{1/2} = 1.253$ and a standard deviation of $[(\pi/2)(\pi - 3)/3]^{1/2} = 0.272$. Many of the early researchers in this area were aware of the possible deficiencies of the R_T/s_T statistic in the presence of data generated by short memory I(0) processes combined with a long memory component. Anis and Lloyd (1976) determine the small sample bias of the R/S statistic; while Mandelbrot (1972, 1975), Mandelbrot and Wallis (1968), Davies and Harte (1987), Aydogan and Booth (1988), and Lo (1991) all discuss the lack of robustness of the R/S statistic in the presence of short term memory and heteroskedasticity. Lo (1991) suggests the modified rescaled range statistic,

$$Q_T = R_T/\sigma_T(q), \tag{50}$$

where

$$\sigma_T^2(q) = c_0 + 2\sum_{j=1}^{q} w_j(q)c_j, \tag{51}$$

c_j is the jth-order sample autocovariance of y_t and $w_j(q)$ are the Bartlett window weights of

$$w_j(q) = 1 - [j/(q+1)] \quad \text{for} \quad q < T.$$

In the context of unit root tests Phillips (1987) shows the consistency of σ_q^2 if $T \to \infty$ and $q \sim O(T^{1/4})$. Lo (1991) shows that in the presence of long memory $T^{-1/2}Q_T$ weakly converges to the range of a Brownian Bridge, the distribution function of which is given by Feller (1951). The distribution function of the range, $F(x)$, given by Kennedy (1976) and Siddiqui (1976) is

$$F(x) = \sum_{j=-\infty}^{\infty} (1 - 4x^2j^2)\exp[-2x^2j^2].$$

The distribution is positively skewed, and Lo (1991) tabulates fractiles of the distribution and shows the modified rescaled range test to be consistent against a fairly general class of long range dependent stationary Gaussian alternatives. However, a major practical difficulty concerns the choice of q and how to distinguish between short range dependencies and long range dependencies. Simulation evidence to be discussed at the end of Section 4 has generally been unfavorable to this approach.

4.2. Unit root tests in the presence of I(d)

Sowell (1990) considers the limiting distribution of the OLS coefficient estimate in an AR(1) model when the true data generating process is $I(1 + d)$. When $d = 0$, the estimate of ϕ reduces to the well-known result derived by Phillips (1987), of

$$T(\hat{\phi} - 1) \Rightarrow (1/2)\{B(1)^2 - 1\} \Big/ \left\{\int_0^1 B(t)^2\,\mathrm{d}t\right\}. \tag{52}$$

However, on assuming that the disturbances u_t are not necessarily normally distributed, but have zero mean and have a finite rth moment for some r such that $r \geqslant \max[4, -8d/(1+2d)]$, Sowell (1990) shows that the asymptotic distribution of the OLS estimate of ϕ only has a nonzero density over the whole real line for the special case of $d = 0$, i.e., a unit root. For other values of d, the asymptotic distribution of $\hat{\phi}$ is a complicated function of two distributions which both depend on fractional Brownian motion. Furthermore, the standard t statistic for a unit root only converges to a well-defined density when $d = 0$.

It is well-known, however, that unit root tests are consistent against I(d) alternatives. A related study by Diebold and Rudebusch (1991b) evaluates the power performance by simulation of the Dickey–Fuller unit root test when the true data generating process is fractionally integrated white noise and AR(1) processes and for sample sizes of $T = 50, 100, 250$. Not surprisingly, the power of the Dickey–Fuller test grows more slowly with divergence of d from one than with the divergence of the AR parameter ϕ from one. Hence the Dickey–Fuller test performs relatively poorly in distinguishing between the I(1) null hypothesis and the I(d) alternative. A related study by Hassler and Wolters (1994) finds the Phillips and Perron unit root test to perform similarly to that of the Augmented Dickey–Fuller test; and with a nonstationary value of $d = 0.75$ generating the fractional white noise, the rejection frequencies of the unit root hypothesis are about 50% when $T = 100$ and about 70% with $T = 250$.

Lee and Schmidt (1996) consider the performance of the KPSS test of Kwiatkowski, Phillips, Schmidt, and Shin (1992) which was originally designed to test an I(0) null hypothesis versus an I(1) alternative. The KPSS test involves taking the residuals e_t from a regression of y_t on an intercept and time trend and forming the partial sum S_t as in (3) of the residuals and to compute the same long run variance formula $\sigma_T^2(q)$ in (51) as by Lo (1991). The KPSS test for stationary is then

$$\eta_\tau = T^{-2} \sum S_t^2 / \sigma_T^2(q), \tag{53}$$

and the KPSS test η_μ is also based on (53) except that the residuals are derived from a regression on an intercept only. Lee and Schmidt (1996) show that the two KPSS tests are both consistent against an I(d) alternative and that the KPSS tests can be used to distinguish short memory from long memory stationary processes. Lee and Schmidt (1996) show that under the I(d) alternative hypothesis the KPSS test statistics converge to functions of fractional Brownian motions which are relatively natural extensions of the second level Brownian bridges previously defined by MacNeill (1978) and Schmidt and Phillips (1992). Lee and Schmidt (1996) also include some Monte Carlo evidence and conclude that the KPSS test has power properties similar to the adjusted rescaled range statistic of Lo (1991) in distinguishing I(0) from I(d) behavior. Robinson (1991) derives the Lagrange Multiplier (LM) test for fractional white noise in the disturbances of a linear regression under the standard assumptions. In particular, Robinson (1991) shows that the LM test for H_0: $d = 0$ versus H_1: $y_t = \beta' x_t + u_t$, $(1 - L)^d u_t = \varepsilon_t$ with $0 \leqslant d \leqslant 0.5$, is given by the statistic

$$LM_1 = T\left[\sum_{j=1}^{T-1} j^{-1} r_j\right]^2 \Big/ (\pi^2/6). \tag{54}$$

where r_j is the jth-order sample autocorrelation coefficient of the OLS residuals. Under the null, LM_1 will have an asymptotic chi-squared distribution with one degree of freedom. Interestingly, Robinson (1991) also shows that to a first-order approximation the same test statistic results from the alternative hypothesis that the process is self-similar, with autocovariance function given by Eq. (11).

Wu (1992) considers the related issue of testing the unit root hypothesis versus the one-sided alternatives of $d < 1$ or $d > 1$. The tests are modified one-sided locally best invariant (LBI) alternatives based on the technique developed by King and Hillier (1985). Wu (1992) shows that the LBI test for H_0: $d = 0$ versus H_1: $d > 0$ in the fractional white noise model (13) is given by

$$LM_2 = -2 \sum_{j=1}^{T-1} j^{-1} r_j,$$

A test of H_0: $d = 0$ versus H_2: $d < 0$ requires using $-LM_2$. Wu (1992) and Agiakloglou and Newbold (1994) also consider forms of the LM statistic to test the same hypothesis and generalize the test statistics to deal with ARFIMA processes under the alternative hypotheses. Beran (1992b) has discussed alternative tests for long range dependence.

Although not specifically dealing with long memory, Blough (1992) and Faust (1994) discuss the near observational equivalence of difference stationary and trend stationary processes. These articles highlight the general difficulty of distinguishing between competing models for the low frequency components of series, and to this extent provide an additional motivation for the class of $I(d)$ processes.

4.3. *Regression with I(d) disturbances*

Taqqu (1975) and Yajima (1988) consider estimation of regression parameters in the presence of disturbances exhibiting long memory. A special case arises in the estimation of the population mean of a long memory process with autocovariance function, $\gamma_k \approx ck^{2d-1}$, as given by (6), (8), or (12). Taqqu (1975) derives the well-known result concerning the properties of the OLS estimate of the mean or intercept parameter μ in the model $y_t = \mu + u_t$, where u_t has an autocovariance function of $\gamma_k \approx ck^{2d-1}$. Taqqu (1975) shows that the sample mean converges at a rate of $T^{1/2-d}$ to an unspecified limiting distribution. Furthermore, the estimator of the uncorrected sample standard deviation,

$$s_T = \left\{T^{-1} \sum (y_t - \bar{y})^2\right\}^{1/2}, \tag{55}$$

and on appropriately normalizing,

$$cT^{-2d}\{s_T - \mathrm{E}(s_T)\} \Rightarrow R, \tag{56}$$

so that weak convergence occurs to the Rosenblatt process denoted by R, which is expressed in terms of Wiener-Ito-Dobrushin integrals. Taqqu (1975) also shows that the sample mean, $\bar{y}$, and s_T are not independent. The sample mean has an asymptotic variance of

$$T^{1-2d}\operatorname{var}(\bar{y}) = c_2/[d(2d-1)], \tag{57}$$

where c_2 is another constant. Also,

$$cT^{1-2d}\{\bar{y} - \mathrm{E}(\bar{y})\}^2 \Rightarrow [d(2d-1)]^{-1}\chi_1^2, \tag{58}$$

where χ_1^2 denotes a chi-squared random variable with one degree of freedom. Adenstedt (1974) shows that the loss in asymptotic efficiency from use of the sample mean as opposed to estimating μ from Generalized Least Squares (GLS) is surprisingly small. There is only a loss of 2% in efficiency when estimating in intercept in the presence of a stationary and invertible disturbance. Similarly, Yajima (1988) considers the efficiency of the sample mean, i.e., the OLS and the GLS estimator with known covariance matrix. Some extensions of the above results are provided by Samarov and Taqqu (1988) and Yajima (1985, 1988), who have considered OLS and GLS in the context of the regression model

$$y_t = \beta' x_t + u_t,$$

where u_t is a long memory process and x_t contains polynomial functions of time. The above articles find expressions for the asymptotic efficiency loss from using OLS, rather than GLS. As in the case of the regression with just an intercept, the loss of efficiency associated with OLS is not necessarily severe. Robinson (1990) extends this investigation to the case where x_t also contains stochastic regressors. The resulting parameter estimates appear to converge at different rates and Robinson (1990) notes that in certain cases, singular limiting distributions may result. It should also be noted that Carlin, Dempster, and Jonas (1985) and Carlin and Dempster (1989) have suggested a Bayesian estimator in the context of long memory models and the Hurst coefficient.

Cheung and Lai (1993) have considered testing for the fractional cointegration of two time series, y_t and x_t, which are both I(d) but are fractionally cointegrated, i.e., CI(d,b). Hence $\varepsilon_t = y_t - \beta x_t$ and ε_t is I($d-b$), where $(d-b) > 0$. In the case of $(d-b) > 1/2$, Cheung and Lai (1993) show that the OLS estimator of β converges in probability to zero for all $\delta > 0$, such that

$$T^{b-\delta}(\hat{\beta} - \beta) = \left[T^{2d-b+\delta}\left(\sum x_t\varepsilon_t\right)\right]\left[T^{-2d}\sum x_t^2\right]^{-1}$$

and, when $1/2 > (d-b) \geqslant 0$, then $T^{2d-b}(\sum x_t\varepsilon_t)$ converges to a function of Brownian motion. This motivates a test for fractional cointegration to be based on the OLS residuals. Cheung and Lai (1993) apply the GPH estimator to the OLS residuals having simulated the power of the procedure on residuals. An interesting topic for future research is to efficiently estimate the general error correction model associated with fractionally integrated series.

4.4. Distribution of sample autocorrelations from an I(d) process

Brockwell and Davis (1987) consider the asymptotic distribution of sample autocorrelations from an intermediate I(d) process with $d \in (-0.5, 0)$ and innovations not necessarily Gaussian. On denoting the sample autocorrelation at lag k as r_k, which is an estimator of ρ_k, the corresponding population autocorrelation, then Brockwell and Davis (1987) show

$$T^{1/2}(r_k - \rho_k) \to \mathrm{N}(0, V_k),$$

and V_k is derived from the asymptotic covariance matrix, using the usual formula of Bartlett.

Hosking (1984) shows the above result to be valid for $d \in (-0.5, 0.25)$. However, outside this range of values of d the asymptotic distribution of the estimators of the autocorrelations depends on the range of values of d, the fractional differencing parameter. In particular,

$$\{T/\log(T)\}^{1/2}(r_k - \rho_k) \to \mathrm{N}[0, V_k(d)] \quad \text{for} \quad d = 0.25$$

and

$$T^{1/2-d}(r_k - \rho_k) \to D \qquad \text{for} \quad 0.25 < d < 0.50,$$

where D is a nonstandard distribution. Hence the rate of convergence is slower than the conventional rate in this range of values of d. Hosking also shows that

$$T^{1/2}[(r_k - \rho_k)(1 - \rho_k)^{-1} - (r_j - \rho_j)(1 - \rho_j)^{-1}] \tag{59}$$

does converge to a nondegenerate normal distribution and permits the possibility of conventional estimation of the suitably weighted autocorrelations. Newbold and Agiakloglou (1993) derive some results on the bias of estimated autocorrelations from fractional processes.

4.5. Semiparametric estimation of d in the frequency domain

Geweke and Porter-Hudak (1983), henceforth GPH, suggested a semiparametric estimator of the fractional differencing estimator, d, that is based on a regression of the ordinates of the log spectral density on trigonometric function. The estimator exploits the theory of linear filters to write the process $(1 - L)^d y_t = u_t$, where $u_t \sim \mathrm{I}(0)$, as

$$f(\omega)_y = |1 - \mathrm{e}^{-i\omega}|^{-2d} f(\omega)_u, \tag{60}$$

where $f(\omega)_y$ and $f(\omega)_u$ are the spectral densities of y_t and u_t respectively. Then (60) can be expressed as

$$\log\{f(\omega)_y\} = \{4\sin^2(\omega/2)\}^{-d} + \log\{f(\omega)_u\},$$

$$\log\{f_y(\omega_j)\} = \log\{f_u(0)\} - d\log\{4\sin^2(\omega_j/2)\} + \log[f_u(\omega_j)/f_u(0)]. \tag{61}$$

GPH suggest estimating d from a regression based on (61) using spectral ordinates $\omega_1, \omega_2, \ldots, \omega_m$, from the periodogram of y_t, that is $I_y(\omega_j)$. Hence, for $j = 1, 2, \ldots, m$,

$$\log\{I_y(\omega_j)\} = a + b\log\{4\sin^2(\omega_j/2)\} + v_j, \tag{62}$$

where

$$v_j = \log[f_u(\omega_j)/f_u(0)] \tag{63}$$

and v_j is assumed to be i.i.d. with zero mean and variance $\pi^2/6$. When u_t is white noise, ε_t, then the regression (62) should provide a good estimate of d. When u_t is autocorrelated, GPH show that (62) holds approximately for frequencies in the neighborhood of zero. If this neighborhood shrinks at an appropriate rate with sample size, then the GPH procedure should realize a consistent estimator of d. If the number of ordinates m is chosen such that $m = g(T)$, where $g(T)$ is such that $\lim_{T\to\infty} g(T) = \infty$, $\lim_{T\to g}\{g(T)/T\} = 0$, $\lim_{T\to g}\{(\log(T)^2)/g(T)\} = 0$, then the OLS estimator of d in (62) will have the limiting distribution

$$(\hat{d}_{GPH} - d)/\{\operatorname{var}(\hat{d}_{GPH})\}^{1/2} \Rightarrow N(0,1).$$

The $\operatorname{var}(\hat{d}_{GPH})$ is obtained from the usual OLS regression formula, either using the regression residual variance or alternatively setting it as $\pi^2/6$. It is clear from this result that the GPH estimator is not $T^{1/2}$ consistent and will converge at a slower rate. Geweke and Porter-Hudak (1983) are able to prove consistency and asymptotic normality only for $d < 0$, while Robinson (1990) provides a proof of consistency for $0 < d < 0.50$; also see Kunsch (1986).

A major issue in the application of the GHP estimator has been the choice of m when u_t is autocorrelated. Diebold and Rudebusch (1989) typically choose $m = T^{1/2}$, while Sowell (1992b) has argued that m should be based on the shortest cycle associated with long-run behavior. For example, with 40 years of data and the a priori view that 2 years is the shortest cycle, then m would be chosen as $40/2 = 20$ ordinates. This decision rule is deliberately independent of the sampling frequency of the date since, for example, with quarterly data, m would also be selected as $160/8 = 20$. Another possibility is to choose m such that the regression residual variance is approximately equal to $\pi^2/6$.

While the GPH estimator is simple to apply and is potentially robust to nonnormality, the behavior of $\hat{d}_{GPH}$ in the presence of substantial autocorrelation of u_t reduces its potential attractiveness. In particular, Agiakloglou, Newbold, and Wohar (1992) show it possesses 'serious bias' and is very inefficient when u_t is AR(1) or MA(1) and the AR or MA parameter is quite large. Also, if an investigator wishes to obtain estimates of short run ARMA parameters as well as d, then filtering the original series by the operator, $(1 - L)^d$ where d is replaced with $\hat{d}_{GPH}$, and estimating the ARMA parameters from the filtered series will provide two-step estimates with a currently unknown sampling distribution. Most applications of this procedure, such as Diebold and

Rudebusch (1989, 1991b), typically assume normality of the filtered series and use approximate MLE to obtain estimates of the ARMA parameters. Even if a quasi-MLE is used in this situation, appropriate inference is likely to remain a difficulty. Hassler (1994) has carried out a simulation study of a variant of the GPH procedure applied to the seasonal ARFISMA process of Porter-Hudak (1990) in (44). The results generally indicate deficiencies with the semiparametric regression. Hurvich and Ray (1995) have considered the bias of the GPH estimator in the case when the true data generating process is a nonstationary ARFIMA process with a value of $d > 0.5$.

In a series of papers Robinson considers various frequency domain approaches to estimating the long-range dependency parameter. These papers are concerned with finding consistent estimates of the Hurst coefficient, or equivalently fractional d, in the absence of any parameterization of the autocovariance function. Robinson (1992) considers the properties of a discretely averaged periodogram,

$$F(\omega) = \int f(\lambda)\,\mathrm{d}\lambda,$$

where the averaging is over the neighborhood $\omega \in (0, \lambda)$. Robinson (1992) shows that $F(\omega)$ converges in probability to one for a sequence λ which tends to zero more slowly than $1/T$, as $T \to \infty$. For any slowly varying function $\Xi(.)$, then

$$F(q\omega)/F(\omega) \approx q^{2(H-1)}\{\Xi(1/q\omega)/\Xi(1/\omega)\}, \tag{64}$$

$$\approx q^{2(H-1)}. \tag{65}$$

Then as $\omega \to 0+$ and Robinson (1992) establishes consistency of the estimator,

$$\hat{H}_q = 1 - \{2\log(q)\}^{-1}\log\{F(q\omega_m)/F(\omega_m)\}, \tag{66}$$

so that $\hat{H}_q \to H$ as $T \to \infty$, and where q is a chosen scalar such that $0 < q < 1$ and $\omega_1, \omega_2, \dots, \omega_m$ are the frequencies of the periodogram used in estimation. Lobato and Robinson (1996) are able to establish the limiting distribution of $\hat{H}_q$ after assuming normality of the y_t process. For $0 < d < 1/4$, i.e., $1/2 < H < 3/4$, the estimator is $m^{1/2}$ consistent, where m is the number of ordinates of the periodogram used in estimation. Then $m^{1/2}(\hat{H}_q - H)$ converges to a limiting normal distribution, and for any H and a bandwidth number m, then an optimal value q exists to minimize estimation MSE. However, as previously mentioned in the context of the asymptotic distribution of sample autocorrelations, there is a discontinuity at $H = 3/4$, or $d = 1/4$, in the asymptotic distribution theory. For $3/4 < H < 1$, H_q converges at rate m^{1-2d}, i.e., $m^{2(1-H)}$ to a nonnormal distribution. Lobato and Robinson are able to establish the properties of the $\hat{H}_q$ estimator under quite weak assumptions regarding the slowly varying function $\Xi(.)$.

Despite the amount of theoretical work in attempting to devise robust semiparametric estimators of the long memory parameter, there is substantial evidence documenting their poor performance in terms of bias and mean squared error. See Agiakloglou, Newbold, and Wohar (1992), Janacek (1982), and Hurvich and Beltrano (1994) for the GPH estimator, Lee and Schmidt (1996), Chen, Abraham, and Peiris (1994), Cheung (1993b), Choi and Wohar (1992), Hassler (1993, 1994), Hauser (1994), and Reisen (1994), who examine a variety of R/S statistics, trimmed periodogram versions of GPH, and related estimators. Hurvich and Ray (1995) consider the bias in the GPH estimator when $d > 0.5$. Overall the consensus of evidence is somewhat negative about semiparametric estimation, with adjustments to the periodogram at low frequencies appearing unlikely to radically improve their small sample performance.

4.6. Semiparametric estimation of d in the time domain

An alternative to periodogram based estimation is to directly use the sample autocorrelations. Robinson (1990) considers such an estimate based upon high lags of γ_k, but notes that such a procedure has the disadvantage of being based on the requirements that the γ_k are always eventually positive. An alternative GMM estimator, or Minimum Distance Estimator (MDE) based on sample autocovariances, has been studied by Tieslau, Schmidt, and Baillie (1995), henceforth TSB. Their estimator uses blocks of n sample autocovariances, $\gamma_k, \gamma_{k+1}, \ldots, \gamma_{k+n}$. From using the results of Hosking (1984), who derives the asymptotic distribution of sample autocovariances from fractional white noise, TSB evaluate the asymptotic efficiency for their MDE. As noted by Hosking (1984) and formally proved by Dahlhaus (1988, 1989), the unusual behavior of the score vector makes the rate of convergence to the limiting distribution dependent on the value of d. For $-0.5 < d < 0.25$, the MDE studied by TSB converges to a limiting normal distribution at the conventional $T^{1/2}$ rate, while for $d = 0.25$, the MDE still converges to a normal distribution, but at a rate of $\{T/\log(T)\}^{1/2}$. For $0.25 < d < 0.50$, the MDE converges at rate of $T^{1/2-d}$ to a nonstandard distribution. TSB then evaluate the theoretical asymptotic efficiency of the MDE for various parameter values d in the range $-0.50 < d < 0.25$ and for different choices of k and n. The efficiency loss for increasing k is found to be quite large and the estimator is not generally very promising.

The results of Hosking (1984) in (59) suggest that it may be possible to obtain a semiparametric estimate of d in the time domain that converges at the standard rate, for the case where $0.25 \leqslant d < 0.50$. However, some simulation evidence indicates that the functional form of (59) may not be very helpful and that an estimator based upon differenced autocorrelations will be quite inefficient.

The potentially interesting application of semiparametric estimation is to ARFIMA models with substantial short memory dynamics. If γ_k are the autocovariances of an ARFIMA(0, d, 0) process and γ_k^* are the autocovariances of an ARFIMA(p, d, q) process, then in the limit

$$\gamma_k^* \approx [\theta(1)/\phi(1)]\gamma_k, \tag{67}$$

so that the rate of decay depends only on the differencing parameter d, and not on the short-run dynamics. This suggests that an estimate of d based upon ratios of autocorrelations would be a useful estimator.

4.7. Maximum likelihood estimation

Several authors have considered joint estimation of the parameters in the ARFIMA(p, d, q) model (25) under the assumption of normality. The $(p + q + 3)$-dimensional vector of parameters is $\lambda' = (\mu \beta')$, where

$$\beta' = (\mathrm{d}\ \phi_1 \ldots \phi_p\ \theta_1 \ldots \theta_q\ \sigma^2).$$

Li and McLeod (1986) consider the asymptotic properties of the MLE in the case of the intercept μ being either known or zero. They assert that

$$T^{1/2}\{(\hat{\beta} - \beta)\} \to \mathrm{N}\left[0, \lim_{T \to \infty} \{I(\beta)/T\}^{-1}\right], \tag{68}$$

and that with known intercept μ, the vector of remaining parameter estimates, $\hat{\beta}$, will be $T^{1/2}$ consistent and will converge to a limiting normal distribution. The form of the information matrix is given by

$$I(\beta) = \begin{bmatrix} I_{p,q} & J \\ J' & \pi^2/6 \end{bmatrix}, \tag{69}$$

where $I_{p,q}$ is the usual information matrix of the ARMA parameters and

$$J' = [\gamma_0^{ud}\ \gamma_1^{ud} \ldots \gamma_{p-1}^{ud}\ \gamma_0^{vd}\ \gamma_1^{vd} \ldots \gamma_{q-1}^{vd}],$$

where

$$\gamma_j^{ud} = \sum_{i=0}^{\infty} (j + i + 1)^{-1} c_i,$$

$$\gamma_j^{vd} = \sum_{i=0}^{\infty} (j + i + 1)^{-1} b_i,$$

$$\phi(L)^{-1} = \sum_{i=0}^{\infty} c_i L^i,$$

$$\theta(L)^{-1} = \sum_{i=0}^{\infty} b_i L^i.$$

For the fractional white noise process, $(1 - L)^d y_t = \varepsilon_t$, $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$, then (68) reduces to the well-known result that

$$T^{1/2}(\hat{d} - d) \Rightarrow \mathrm{N}(0, 6/\pi^2). \tag{70}$$

The above is an example of the general and rather surprising property of ARFIMA models that the asymptotic variance of their parameter estimates are independent of the value of d. This is in contrast to ARMA models, where each parameter generally occurs in at least one element of the information matrix. The form of $I(\beta)$ in (69) also implies that the MLE of d will generally by asymptotically correlated with the ARMA parameter estimates. While Galbraith and Galbraith (1974) and Newbold (1974) provide parametric results for the inverse of $I(\lambda)$ for the class of stationary and invertible ARMA models, no corresponding results are yet available for the ARFIMA process.

On using Whittle's (1951) approach of approximating the exact likelihood in the frequency domain, the (j,k)th element of $I(\lambda)^{-1}$ is given by

$$I_{jk}(\lambda) = (1/4\pi) \int_{-\pi}^{\pi} [(2\pi/\sigma^2)\{\delta \log f(\omega|\lambda)/\delta\lambda_j\}]$$
$$\times [(2\pi/\sigma^2)\{\delta \log f(\omega|\lambda)/\delta\lambda_k\}]' \, d\omega. \tag{71}$$

For the case of μ unknown, the formal proof of asymptotic normality and the appropriate rates of convergence of the MLE for the ARFIMA(p,d,q) process is due to Dahlhaus (1988, 1989) for the $0 < d < 0.5$ case and to Moehring (1990) for the case of $-0.5 < d < 0$. Joint estimation of the parameter vector λ by MLE will give a limiting distribution of

$$D_T(\hat{\lambda} - \lambda) \Rightarrow \mathrm{N}\{0, [D_T^{-1} I(\lambda) D_T^{-1}]^{-1}\}, \tag{72}$$

where

$$\mathrm{diag}\{D_T\} = [T^{1/2-d}, T^{1/2}, \dots, T^{1/2}]. \tag{73}$$

Hence the MLE of μ will converge at the slow rate of $T^{1/2-d}$, while all the other parameter estimates converge at the standard $T^{1/2}$ rate.

Sowell (1986, 1992a) derives the exact MLE of the ARFIMA process with unconditional Normally distributed disturbances ε_t. The log-likelihood is then

$$\mathscr{L} = -(T/2)\log(2\pi) - (1/2)\log|\Omega| - (1/2)Y'\Omega^{-1}Y, \tag{74}$$

where $\{\Omega\}_{ij} = \gamma_{|i-j|}$ and Y represents a T-dimensional vector of the observations on the process y_t. While Sowell's (1992a) full MLE is theoretically appealing, it is computationally demanding since it requires the inversion of a $T \times T$ matrix of nonlinear functions of the hypergeometric function at each iteration of the maximization of the likelihood. The method requires all the roots of the autoregressive polynomial to be distinct and for the theoretical mean parameter μ to be either zero or known.

There are several alternative approximate MLE of the ARFIMA(p,d,q) model in (25) under normality. Whittle (1951) notes that the autocovariance matrix Ω can be diagonalized by transforming the vector Y into the frequency domain and can approximate the log-likelihood by

$$\mathscr{L} = \sum_{j=1}^{T-1} \log[(2\pi) f(\omega_j)] + \sum_{j=1}^{T-1} [I_T(\omega_j)/f(\omega_j)]. \tag{75}$$

The above approximate MLE has been used by Boes, Davis, and Gupta (1989), who concentrate out σ^2 to obtain

$$\sigma^2(\lambda) = (2\pi/T) \sum_{j=1}^{T-1} [I_T(\omega_j)/f(\omega_j)].$$

Eq. (71) is sometimes known as the 'Whittle Likelihood', and Tschernig (1992) has studied the small sample properties of this version of the MLE by simulation. An alternative frequency domain approximate MLE is due to Fox and Taqqu (1986), which numerically minimize the quantity

$$\sum \{I(\omega_j)\}/f(\omega_j; \theta), \tag{76}$$

where $I(\omega_j)$ is the periodogram evaluated at frequency ω_j and the summation is over m frequencies.

Chung and Baillie (1993) consider a Conditional Sum of Squares (CSS) estimator in the time domain, which is obtained by minimizing the quantity

$$S = (1/2)\log(\sigma^2) + (1/2\sigma^2) \sum_{t=1}^{T} \{\phi(L)\theta(L)^{-1}(1 - L)^d(y_t - \mu)\}^2. \tag{77}$$

Some results concerning the small sample performance of the CSS estimator are reported in Chung and Baillie (1993). They conclude that for the ARFIMA(0,d,0) model, with $T = 100$ and with the mean unknown, CSS is extremely similar to Sowell's full MLE. For the ARFIMA(p,d,q) model with unknown mean and complicated ARMA dynamics, i.e., $p, q > 2$, the CSS estimator can produce substantial biases in samples of 300. The estimation of the intercept μ can substantially affect the properties of the other parameter estimates. However, the CSS estimator performs quite well for ARFIMA models with known mean parameter and $T = 500$. Some further simulation evidence is provided by Cheung and Diebold (1994). Interestingly enough they find that the Fox–Taqqu estimator is preferable to Sowell's full MLE when the mean of the process, μ, is unknown. In their application of the Fox–Taqqu estimator, Cheung and Diebold (1994) essentially use a two-step estimator which replaces the unknown mean parameter μ with the sample mean of y_t before estimating the other parameters with the Fox–Taqqu estimator. Cheung and Diebold (1994) find evidence that their estimator is more satisfactory than the full MLE of Sowell in the sense of bias and MSE. These differences are again brought about by the key fact that any estimator of μ converges at a slower rate than the other parameter estimates which are all $T^{1/2}$ consistent. In some applications (e.g., Sowell, 1992a), the data series are differenced and the process is estimated with $d < -0.50$. This strategy has the advantage of removing the troublesome intercept parameter, and some simulation evidence on the efficacy of the procedure is presented by Smith, Sowell, and Zin (1993). A further time domain approximate MLE is described by Haslett and Raftery (1989), who only consider d in the range of $0 \leqslant d \leqslant 0.5$. While the full MLE is obviously desirable when appropriate, more complicated models can only currently be estimated by approximate MLE, which can be conveniently done by minimizing the CSS function in (77). For example, Baillie, Chung, and Tieslau (1995) estimate an ARFIMA process with a conditional variance process following a GARCH(1, 1) formulation. They estimate monthly inflation series with an ARFIMA(0, d, 1)–GARCH(1, 1) process.

Chung (1996) provides some results on the estimation of the Gegenbauer process in (34) by approximate MLE, using the CSS method. An important finding is that the MLE of ξ converges at a rate of $O_p(T)$ to a function of Brownian motions. The other parameter estimates converge at the usual $T^{1/2}$ rate.

5. Long memory volatility processes

The topic of long memory and persistence has recently attracted considerable attention in terms of the second moment of a process. Many of the obvious examples of long memory processes have emerged in studies of financial market data and will be described further in Section 6 of this article. The desire to develop theoretical tests and models for long memory volatility has been the result of encountering data which strongly exhibit this phenomenon. As with virtually all volatility processes, the choice of model has generally not been dictated by economic or finance theory, but rather mathematical tractability and/or data compatibility. The first contribution in this regard was Taylor (1986), who noticed an apparent stylized fact that the absolute values of stock returns tended to have very slowly decaying autocorrelations. Ding, Granger, and Engle (1993) note the same fact for the powers of daily returns and Dacorogna, Muller, Nagler, Olsen, and Pictet (1993) find similar phenomena for squared exchange rate returns, recorded every twenty minutes over a four-year period.

A long memory conditional variance process can be set up from the same foundations as the ARCH model of Engle (1982). It is natural to define a discrete time, real valued stochastic process ε_t,

$$\varepsilon_t = \xi_t \sigma_t, \tag{78}$$

where ξ_t is i.i.d. with $E(\xi_t) = 0$ and $\mathrm{var}(\xi_t) = 1$. The variable σ_t^2 is a time-varying, positive, and measurable function of the information set at time $t-1$, denoted by Ω_{t-1}, and σ_t^2 is known as an ARCH process. The GARCH(p, q) specification of Bollerslev (1986) is defined as

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2, \tag{79}$$

with $\alpha(L)$ and $\beta(L)$ being polynomials of order q and p in the lag operator. For stability all the roots of $\alpha(L)$ and $\{1 - \alpha(L) - \beta(L)\}$ are constrained to lie outside the unit circle. The GARCH(p,q) process can also expressed as an ARMA(m,p) process in ε_t^2, where $m = \max(p,q)$,

$$\{1 - \alpha(L) - \beta(L)\}\varepsilon_t^2 = \omega + \{1 - \beta(L)\}v_t, \tag{80}$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$ are the 'innovations' in the conditional variance process. When the polynomial $\{1 - \alpha(L) - \beta(L)\}$ in (79) contains a unit root, then the GARCH(p,q) process is a member of the integrated GARCH, or IGARCH(p,q), class of models defined by

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + \{1 - \beta(L)\}v_t,$$

where $\phi(L) = \{1 - \alpha(L) - \beta(L)\}(1 - L)^{-1}$ is of order $m - 1$.

Baillie, Bollerslev, and Mikkelson (1996) have considered a long memory process in the conditional variance, known as Fractionally Integrated Generalized AutoRegressive Conditional Heteroskedasticity, i.e., FIGARCH. This process implies a slow hyperbolic rate of decay for lagged squared innovations and persistent impulse response weights. Also, the cumulative weights tend to zero, a property in common with weakly stationary or stable GARCH processes. However, the impulse response weights of the FIGARCH process decay at a very slow hyperbolic rate. The FIGARCH(p,d,q) process is defined as

$$\phi(L)(1 - L)^d\varepsilon_t^2 = \omega + \{1 - \beta(L)\}v_t, \tag{81}$$

where all the roots of $\phi(L)$ and $\{1 - \beta(L)\}$ lie outside the unit circle. Analogously to (81) the FIGARCH process can also be represented as

$$\{1 - \beta(L)\}\sigma_t^2 = \omega + \{1 - \beta(L) - \phi(L)(1 - L)^d\}\varepsilon_t^2 \tag{82}$$

and as

$$\sigma_t^2 = \omega\{1 - \beta(1)\}^{-1} + \lambda(L)\varepsilon_t^2, \tag{83}$$

where

$$\lambda(L) = \{1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\}. \tag{84}$$

A necessary and sufficient condition for the FIGARCH$(1,d,0)$ process to have nonnegative impulse response coefficients, $\lambda_j \geqslant 0$ for positive integer j is for $0 < d < \beta$. Following Baillie, Bollerslev, and Mikkelsen (1996), the polynomial in the lag operator of the impulse response coefficients is denoted by $\gamma(L)$, where

$$\gamma(L) = \sum_{k=0}^{\infty} \gamma_k L^k,$$

which is directly analogous to corresponding calculations for the mean given by Eq. (30). Then,

$$(1 - L)\varepsilon_t^2 = \omega + \gamma(L)v_t \tag{85}$$

and

$$\gamma(L) = (1 - L)^{1-d}\phi(L)^{-1}\{1 - \beta(L)\}. \tag{86}$$

The impact of past shocks on the volatility process is given by the limit of the cumulative impulse response weights,

$$\gamma(1) = \lim_{k\to\infty} \lambda_k = \sum_{j=0}^{\infty} \gamma_j.$$

If $d = 0$, then σ_t^2 is a stable GARCH(p,q) process and $\gamma(1) = 0$, so that there is a direct analogy with trend stationary or I(0) processes in the mean. If $d = 1$, then $\gamma(1)$ will converge to a nonzero finite constant, so that the process is analogous to an I(1) process in the mean. For the stable GARCH(1, 1) process, $\{1 - (\alpha + \beta)L\}\varepsilon_t^2 = \omega + (1 - \beta L)v_t$, the impulse response weights are

$$\gamma(L) = (1 - L)\{1 - (\alpha + \beta)L\}^{-1}(1 - \beta L),$$

and hence $\gamma_0 = 0$, $\gamma_1 = (\alpha - 1)$, and $\gamma_j = \alpha(\alpha + \beta - 1)(\alpha + \beta)^{j-2}$, for $j > 2$. Then $\lambda_k = \alpha(\alpha + \beta)^{k-1}$, so that $\gamma(1) = 0$ and the cumulative response weights are zero in the limit. For the IGARCH(1, 1) process, $\lambda_k = (1 - \beta)$ for all lags $k > 1$ and the cumulative impulse response weights tend to a nonzero constant $\gamma(1) = 1 - \beta$. For the FIGARCH process and for a value of $d > 1$, then $\gamma(1)$ will be infinite, while for the FIGARCH $(1, d, 0)$ process,

$$\lambda_k = [\Gamma(k + d - 1)/\{\Gamma(k)\Gamma(d)\}][(1 - \beta) - (1 - d)/k]. \tag{87}$$

The cumulative effect of a shock will be zero on the volatility process since $\gamma(1) = 0$; and from Stirling's approximation,

$$\lambda_k \approx [(1 - \beta)/\Gamma(d)]k^{d-1}, \tag{88}$$

so that hyperbolic decay occurs in the response of the conditional variance to past shocks. Since $\lambda(1) = 1$, it follows that $\mathrm{E}(\varepsilon_t^2)$ is undefined, and hence the second moment of the unconditional density of ε_t is infinite. The FIGARCH process is clearly not weakly stationary, a feature it shares with the IGARCH process. Approximate maximum likelihood estimates of the parameters of the FIGARCH(p, d, q) process in (81) can be obtained by maximizing the Quasi Maximum Likelihood, which realizes $T^{1/2}$ consistent estimates of the FIGARCH parameters. Then,

$$T^{1/2}(\hat{\theta}_T - \theta_0) \Rightarrow \mathrm{N}\{0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}\}, \tag{89}$$

where $A(.)$ and $B(.)$ represent the Hessian and outer product gradient, respectively, and θ_0 denotes the true parameter values. Simulation evidence indicates that the limiting distribution theory works well in sample sizes of 1500 and 3000. Baillie, Bollerslev, and Mikkelson (1996) also report the effects of estimating stable GARCH processes where the true data generating process is FIGARCH. The sum of the estimated GARCH(1, 1) parameters is always close to one, which implies integrated GARCH, or IGARCH, behavior and suggests that the apparent widespread IGARCH property so frequently found in high frequency asset pricing data (see Bollerslev, Chou, and Kroner, 1992) may well be spurious, that the IGARCH process is a poor diagnostic at distinguishing between integrated, as opposed to long memory, formulations of the conditional variance process.

Bollerslev and Mikkelson (1996) extend the FIGARCH process to FIEGARCH, to correspond with Nelson's (1991) Exponential ARCH model to allow for nonsymmetries. The FIEGARCH(p,d,q) model is then

$$\log(\sigma_t^2) = \omega + \phi(L)^{-1}(1-L)^{-d}[1-\lambda(L)]g(\xi_{t-1}), \tag{90}$$

where

$$g(\xi_t) = \theta\xi_t + \gamma[|\xi_t| - \mathrm{E}|\xi_t|] \tag{91}$$

and all the roots of $\phi(L)$ and $\lambda(L)$ lie outside the unit circle. When $d = 0$, the FIEGARCH(p,d,q) process reduces to Nelson's EGARCH process, and when $d = 1$, the process becomes integrated EGARCH. Bollerslev and Mikkelsen (1996) present evidence on the efficacy of QMLE applied to estimate the parameters of the FIEGARCH process and illustrate its application to the pricing of options.

Another route for the modeling of persistence in variance is through the stochastic volatility process developed by Breidt, Crato, and de Lima (1993) and Harvey (1993). The model is then

$$y_t = \xi_t\sigma_t,$$

and

$$\sigma_t^2 = \sigma^2\exp(h_t), \tag{92}$$

where ξ_t is NID(0, 1). In previous work on stochastic volatility models it is commonly assumed that h_t is an AR(1) process, which implies an ARMA(1, 1) representation for $\log(y_t^2)$. If it is assumed that h_t is the fractional white noise process,

$$(1-L)^d h_t = \varepsilon_t, \tag{93}$$

where $\varepsilon_t \sim \mathrm{NID}(0,\sigma_t^2)$, then (92) and (93) generate a long memory stochastic volatility process. Estimation of regular stochastic volatility models has generally been through the state space representation and used QML estimation via the Kalman filter. Since a state space representation does not exist for long memory processes, estimation of the long memory stochastic volatility process is correspondingly difficult. Breidt, Crato, and de Lima (1993) use frequency domain approximate MLE to estimate an ARFIMA$(0, d, 1)$ model for $\log(y_t^2)$, while Harvey (1993) uses the GPH estimator to obtain an estimate of d in a fractional white noise model for $\log(y_t^2)$. The comparison of long memory ARCH and stochastic volatility models remains an interesting area for future research.

6. Applications

6.1. Applications in geophysical sciences

The initial work on long memory processes by Hurst (1951) was concerned with river flow data; and in subsequent work, reported in Hurst (1956), a further 900 geophysical data series of length varying between 40 and 200 years were analyzed by means of the rescaled range statistic. The R/S analysis found the mean value of H to be 0.73 with a standard deviation of 0.08 across the 900 series. Hurst took this as important evidence for the preponderance of the long memory characteristic in such data. A typical series analyzed by Hurst (1951, 1956) and Mandelbrot (1972) is the tree ring data from Mount Campito, which was previously described in Section 2.3 and plotted in Fig. 1. The very slowly decaying autocorrelations in Fig. 2 strongly suggest long memory, and the simplest possible model is the ARFIMA$(0, d, 0)$, or fractional white noise model. Table 3a presents the details of this model estimated by approximate MLE by the CSS method. Extraordinarily, this very simple, one-parameter model accounts for all of the obvious dynamics in the conditional mean of the process. The estimated value of d is 0.449, with a standard error of only 0.010, and implies a Hurst coefficient of 0.949. Apart from having very slowly decaying autocorrelation functions, most tree ring series and climatological data exhibit distinct periods of volatility and tranquility, as evidenced by the LM test for ARCH effects on the squared residuals. On estimating a simple ARFIMA $(0, d, 0)$–GARCH$(1, 1)$ process the estimates for $\alpha = 0.033$ (0.002) and $\beta = 0.950$ (0.005) are highly significant with the score test revealing little evidence of further ARCH effects. Baillie and Bollerslev (1992) derive the prediction density of forecasting from an ARMA model with GARCH$(1, 1)$ innovations and show how long horizon predictions are affected by current volatility. The issue is likely to be very similar for the ARFIMA–GARCH model, and suggests that climatic changes or trends that occur during particularly volatile weather patterns should be interpreted differently than changes during more stable weather regimes.

Analysis of additional tree ring series, mud varvs on river floors, high tides, and other geophysical series are provided by Hipel and McLeod (1978) and Noakes et al. (1988). Their statistical analysis is based on the rescaled R/S, and therefore does not provide very precise estimates of the Hurst coefficient. However, their work does provide more evidence of long memory in a variety of these series. While regular Brownian motion can be derived as a stochastic process from physical laws corresponding to the random behavior of a particle suspended in a liquid, there does not appear to be a readily available explanation to justify the occurrence of fractional Brownian motion. However, Figs. 1 and 2 and Table 3a show that the simple fractional white noise model provides a remarkably good representation of the tree ring series.

6.2. Applications in macroeconomics

The principal economic application of long memory models has been to contribute to the long standing debate as to whether real GNP is difference stationary or trend stationary. Adelman (1965) first suggested the use of long memory models in the context of modeling long run cycles in the macroeconomy. Diebold and Rudebusch (1989) use quarterly post World War II US real GNP data, and on applying the GPH approach to differenced data find an estimated value and asymptotic standard error of $d = -0.50$ (0.27). Sowell (1992b) argues this result is due to the misspecification of short-run dynamics, since Diebold and Rudebusch (1989) use $m = 11$ ordinates in the GPH periodogram regression (62). Sowell (1992b) estimates an ARFIMA(3, d, 2) model of first differenced US real GNP from 1947. I through 1989. IV and obtained a value of $d = -0.59$ (0.35). Estimation of the same data set in levels found essentially equivalent results to Sowell's analysis, with the approximate MLE based CSS estimator, when applied to levels, realized an estimated $d = 0.41$ (0.40) for an ARFIMA(3, d, 2) model. While stringent confidence intervals include both $d = 0$ and $d = 1$ and suggest the likelihood is relatively flat, it is also true that a vast amount of the probability mass is well in the interior of the unit interval.

Price series, particularly over long periods of history, also frequently appear to possess persistence and long memory. Fig. 3 plotted the well-known Wheat Price Index of Beveridge (1925), and Table 1 presents the autocorrelations of the series in levels and differences. This data provides an excellent example of the type of long memory features which has the appearance of being non-stationary in levels and yet also appears overdifferenced. For the Beveridge Wheat Price Index this aspect is also apparent in the estimated models, where there is uncertainty as to whether the series should be analyzed in differences or levels. The favored ARFIMA models in both levels and differences are presented in Table 3b. The estimated value of d indicates a relatively flat likelihood in the approximate range of $d \in (0.40, 0.60)$.

Table 3
Estimated long memory models

(a) *Mount Campito tree rings*, data from 3436BC through 1969AD, $T = 5405$

Model: $(1 - L)^d(y_t - \mu) = \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \omega)$

Parameter	Estimate	Standard error
μ	0.48199	0.03186
d	0.44931	0.01045
ω	0.00639	0.00012

Maximized log-likelihood = 5984.386050
Sample skewness of residuals = − 0.58695
Sample kurtosis of residuals = 4.82550
Ljung–Box statistics: $Q(10) = 11.46059$, $Q(25) = 26.06064$

(b) *Beveridge Wheat Price Index*

Model: $(1 - \phi_1 L - \phi_2 L^2)(1 - L)^d(1 - L)(y_t - \mu) = \varepsilon_t$

Parameter	Estimate	Standard error
μ	0.05367	0.00618
d	− 0.56749	0.07308
ϕ_1	0.54786	0.07978
ϕ_2	− 0.31400	0.05038
σ^2	3.95994	0.29152

Maximized log-likelihood = − 777.502274
Ljung–Box statistics: $Q(10) = 8.27492$, $Q(25) = 42.23435$

Model: $(1 - \phi_1 L - \phi_2 L^2)(1 - L)^d(y_t - \mu) = \varepsilon_t$

Parameter	Estimate	Standard error
μ	2.51400	1.77200
d	0.67122	0.04966
ϕ_1	0.37651	0.06507
ϕ_2	− 0.31511	0.05060
σ^2	4.15775	0.30568

Maximized value of the log-likelihood = − 788.627327
Ljung–Box statistics of residuals: $Q(10) = 16.62535$, $Q(25) = 48.89652$

(c) *US Consumer Price Index*:

Model: $(1 - L)^d[\log \Delta(CPI_t) - \mu] = (1 - \theta L)\varepsilon_t, \quad \varepsilon_t \mid \Omega_{t-1} \sim t(0, \sigma_t^2, v), \quad (1 - \beta L)\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2$

Parameter	Estimate	Standard error
d	0.472	0.065
μ	0.306	0.176
θ	− 0.223	0.083
ω	0.0034	0.0017
α	0.094	0.032
β	0.870	0.040
v	7.817	2.304

Maximized value of the log-likelihood = − 467.54
Ljung–Box statistics of residuals: $Q(10) = 15.6$

The autocorrelations of US Consumer Price Index (CPI) inflation on a monthly basis from 1948 are also given in Table 1. Baillie, Chung, and Tieslau (1995) find similar autocorrelation functions for nine other industrialized countries, and find that the application of unit root tests reveals rejection of the I(1) null from the Augmented Dickey–Fuller regressions and also rejection of the I(0) null form the KPSS tests. As shown in Table 3c, an estimated ARFIMA(0, d, 1) model provides a good explanation of the mean behavior of US CPI inflation. Baillie, Chung, and Tieslau (1995) also estimate by approximate MLE various ARFIMA(0, d, 12)–GARCH(1, 1) models for nine other countries' inflation series. They find that all the countries apart from Japan exhibit evidence of I(d) behavior, and they report feedback between the mean and variance of inflation. The long run properties of these models and their response to shocks are quite different from a high-order AR model. Tieslau (1992) provides further details of estimating long memory models to price and monetary series. A related study by Hassler and Wolters (1995) also considers long memory in inflation, but estimates the d parameter by means of the GPH method.

Crato and Rothman (1994a) use Sowell's (1992a) full MLE approach to estimate ARFIMA models for the macroeconomic time series first analyzed by Nelson and Plosser (1982). Apart from some labor market series, they generally conclude in favor of the series being best characterized by difference stationary processes. Diebold and Rudebusch (1991a) and Haubrich (1992) consider the relationship between consumption and income and the so-called Deaton paradox, where consumption appears too smooth for the permanent income hypothesis to hold. Haubrich (1992) finds that if income follows an ARFIMA process, then the observed variance of both consumption and income is consistent with the permanent income hypothesis. Haubrich and Lo (1993) consider the implications of long memory for the business cycle.

6.3. *Applications in asset pricing models*

Conventional wisdom has typically assumed one unit root in the nominal price of an asset. Then, if p_t is the price of the asset, the continuously compounded rate of return $\Delta \log p_t$ can be expected to be stationary and is usually assumed to be uncorrelated so that it is well approximated as a martingale. Various studies, to be discussed below, have tested for long memory in $\Delta \log p_t$. From the conventional asset pricing formula, if p_t is the price of the asset and x_t are the 'fundamentals' in period t, assuming the transversality condition and the absence of bubbles, then

$$p_t = \sum_{j=0,\infty} \xi^j \mathrm{E}_t x_{t+j}, \tag{94}$$

where $0 < \xi < 1$ and is the discount factor. As shown by Campbell and Shiller (1987) and Baillie (1989), rearrangement of (93) reveals

$$\Delta p_t = (p_t - \xi x_t). \tag{95}$$

The interesting point is that if p_t and x_t are both I(1) processes, then Eq. (95) implies a cointegrating relationship between the asset price and the fundamentals. The failure to find the regular form of CI(1, 1) cointegration between prices and fundamentals is then not necessarily interpretable as a rejection of the asset pricing model, or for that matter, the presence of a bubble. An alternative possibility is that a form of CI(1, $1 - d$) cointegration may be apparent, where the residuals from the cointegrating vector are I(d), rather than I(0). Hence a slower response to shocks and a longer time to adjust back to equilibrium are implied by CI(1, $1 - d$) type cointegration. Two examples of this are in the area of international macroeconomics and are discussed in Section 6.5.

Perhaps the most exciting current application of long memory processes has been concerned with the volatility of asset prices. The work of Ding, Granger, and Engle (1993) and others promises an additional stylized fact in asset pricing, and future research will have to be directed at providing a theoretical explanation. They suggest an Asymmetric Power ARCH, or A-PARCH, model to describe the long memory properties encountered in returns data. The model imposes a power transformation on the conditional standard deviation and the asymmetric absolute innovations; but still implies an exponential decay of the volatility process. It is worth noting that all of the long memory volatility papers discussed in Section 5 find substantial evidence of long memory behavior in either the conditional variance or squared returns. In particular, Baillie, Bollerslev, and Mikkelsen (1996) apply the FIGARCH process to exchange rates, Bollerslev and Mikkelsen (1996) apply the FIEGARCH process to stock prices, and Breidt, Crato, and de Lima (1993), Crato and de Lima (1994), and Harvey (1993) find evidence of long memory stochastic volatility in stock returns and exchange rates respectively.

6.4. Applications to stock returns

Greene and Fielitz (1977) and Aydogan and Booth (1988) used the original *R/S* analysis of Hurst (1951) to test for long memory in common stock returns; while Lo (1991) uses the modified rescaled range statistic (50) on returns from value and equal weighted CRSP indices from July 1962 through December 1987. Lo (1991) finds significant results from using the regular rescaled range statistic and insignificant results from the application of his modified rescaled range statistic. Lo attributes the difference in the test statistics to the short-term persistence within the returns series. Lo also reports the finding of a lack of long-range persistence on annual returns from 1872 through 1986.

As mentioned earlier, the notion of at least stationary (if not uncorrelated) returns seems inherently reasonable, and a unit root is therefore to be expected in the price level series. The conventional CI(1, 1) form of cointegration derived from (94) and (95) in the context of stock prices and dividends have been tested by Campbell and Shiller (1987) and Diba and Grossman (1988). Also see Kaen and Roseman (1986) for a discussion of long memory in asset markets.

6.5. Applications to exchange rates

There is widespread evidence that the logarithm of nominal bilateral exchange rates contain a unit root and, furthermore, that the approximate rate of return is uncorrelated, indicating the appropriateness of the martingale model (e.g., Meese and Singleton, 1981; Baillie and Bollerslev, 1989). An early study by Booth, Kaen, and Koveos (1982) applies the basic rescaled range statistic defined in (47) and (48) to exchange rates. Cheung (1993a), taking monthly data from January 1974 through December 1989, found some evidence for long memory in the French franc/US dollar nominal exchange rates and some marginal evidence for the UK pound/US dollar rate, but no apparent departure from martingale behavior for the German mark, Swiss franc, or Japanese yen. Cheung's work is really the only recorded possible departure from martingale behavior of the nominal exchange rate.

A major issue concerned the speed of adjustment to shocks from disequilibrium. Baillie and Bollerslev (1989) find that, while seven nominal spot exchange rates contained unit roots in their univariate time series representations, they also appeared to be tied together through one cointegrating vector. Subsequent studies by Hakkio and Rush (1991) and Sephton and Larsen (1991) find mixed evidence for the existence of a cointegrating relationship between this set of exchange rates. Diebold, Gardeazabal, and Yilmaz (1994), using the same daily exchange rates over a five-year period as Baillie and Bollerslev (1989), note that the application of the Johansen procedure to test for the number of cointegrating vectors was sensitive to whether or not an intercept was included in the vector autoregression. Diebold, Gardeazabal, and Yilmaz (1994) conclude against the finding of cointegration between the spot exchange rates. Subsequently Baillie and Bollerslev (1994a) find evidence that a linear combination of the same spot exchange rates contains long-range dependence. They estimate d as 0.89 in a fractional white noise model, with an asymptotic standard error of 0.02, and note apparent evidence for the Gegenbauer model in (34) to describe the linear combination of exchange rates.

Many previous studies have been concerned with the properties of real exchange rates and the possible validity of Purchasing Power Parity (PPP) as a long-run phenomenon. In general there has been little support for long-run PPP using data from the present float since 1973. Kim (1990), using the

Johansen test, finds some evidence for cointegration between nominal exchange rates and relative prices with data back to 1910. Diebold, Husted, and Rush (1991) use the approximate MLE method of Fox and Taqqu (1986) to estimate ARFIMA models for annual real exchange rate data from the mid-nineteenth century until the present time. Their results are very supportive of the long-run PPP doctrine, with shocks taking a long but finite time to return to equilibrium. Some supporting evidence is provided by Cheung and Lai (1993), who test for fractional cointegration between the nominal exchange rate and relative prices for annual data from 1914 through 1972. The limitation of data from the recent float clearly presents a difficulty in distinguishing between unit root and fractional behavior. This point has been emphasized in the context of unit root tests by Shiller and Perron (1985) and by Hakkio and Rush (1991) in the context of standard CI(1, 1) tests. However, Crato and Rothman (1994b), estimating ARFIMA models by full MLE, find evidence of mean reversion in UK real exchange rates. In general, the PPP literature is one of the best examples of how many researchers have been misled by the low power of unit root testing procedures and have tended to abandon PPP without sufficient attention to the econometric procedures. Supporting evidence of PPP in a unit root framework has recently been found by Steigerwald (1994).

The monetary model of exchange rate determination implies that if the spot rate and the fundamentals are both I(1), then from (94) and (95) the exchange rate should be cointegrated with the fundamentals. Baillie and Pecchenino (1991) fail to find cointegration between the UK pound/US dollar exchange rate and the fundamentals in terms of relative money supplies and real incomes. However, Baillie and Pecchenino only test for cointegration of the CI(1, 1) form through use of the Johansen test and, when using the GPH estimator, note that it is hard to distinguish whether the real exchange rate has a unit root or is fractionally integrated. A potentially interesting topic for future research is whether the nominal exchange rate is fractionally cointegrated with its fundamentals, in which case disequilibrium shocks would generate very slow decay when returning to equilibrium.

There is widespread statistical evidence that spot and forward exchange rates contain one unit root and are approximately uncorrelated. This is consistent with economic reasoning that returns should be I(0); and it is natural to consider spot and forward rates to be cointegrated, although there is no compelling reason to require them to be CI(1, 1). While many authors have concluded that the forward premium $(s_t - f_t)$ is stationary, others such as Evans and Lewis (1993) have been concerned about possible nonstationarity of the forward premium. Baillie and Bollerslev (1994b) found the forward premium $(f_t - s_t)$ for monthly data from January 1974 through December 1991 to be well described by an ARFIMA$(2, d, 0)$ model. A corollary of this finding is that the forward market forecast error $(s_{t+1} - f_t)$ is also $I(d)$. One interesting implication of this is that the commonly employed 'speculative efficiency' test of Fama (1984),

$$(s_{t+1} - s_t) = \alpha + \beta(f_t - s_t) + \varepsilon_{t+1}, \tag{96}$$

may be inappropriate for testing $\alpha = 0$, $\beta = 1$, and ε_{t+1} uncorrelated, since it suffers from the spurious regression type critique. If such a regression is to be run, then it is also necessary to include the error correction term for fractional cointegration discussed by Granger (1981, 1983). The apparent persistence of the forward premium may also be related to the long memory of the volatility process through an intertemporal asset pricing model as described by Engel (1991). Then, in a risk-neutral environment,

$$f_t - s_t = 0.5\,\mathrm{var}_t(s_{t+1}) - \mathrm{cov}_t(s_{t+1} p_{t+1}), \tag{97}$$

where p_t is the logarithm of domestic prices, so that the variance of the future exchange rate will dominate the risk premium, as any covariation between the spot rate and the price level will be very small due to the smoothness of prices. For risk-averse investors, application of an intertemporal asset pricing models, where investors maximize expected utility subject to successive budget constraints, gives

$$f_t - s_t = 0.5\,\mathrm{var}_t(s_{t+1}) + \mathrm{cov}_t(s_{t+1} q_{t+1}), \tag{98}$$

so the forward premium and also the risk premium is related to the conditional variance of the future spot exchange rate and the conditional covariance of the spot rate with the intertemporal marginal rate of substitution, q_{t+1}. If the spot rate is generated by martingale–FIGARCH process, then a form of long memory would be generated into the forward premium. The implication for models of pricing risk remains an interesting area for future research.

6.6. Applications to interest rates

Shea (1991) applies the GPH procedure to estimate fractional processes on a set of interest rates and discusses the implication of long memory on the variance bounds tests resulting from the term structure. Some of the initial work using the GPH estimator appeared to find long memory in the spreads and some interest rates in levels. Backus and Zin (1993) assume various time series processes, including AR(1), unit root, and fractional white noise, for the short rate returns. They find evidence of long memory and discuss the implication of the presence of fractional integration in the context of the term structure. On comparing the implied forward rates and corresponding yields on maturities of n-period bonds, they conclude that the long memory assumption compares favorably with the alternatives. The estimation of various ARFIMA models to bond series is relatively inconclusive. However, Crato and Rothman (1994a) use full MLE to estimate an ARFIMA$(0,d,1)$ model for annual bond yields from

1900 through 1988 and conclude that d is 0.81 and significantly different from one.

7. Conclusion

The long memory feature of data in some physical sciences has been a well-documented fact for some time. Since the publication in the early 1980s of the work of Granger and Hosking on fractionally integrated processes in discrete time, there has been a steadily increasing interest from economists in this field. For many years the estimation of these models was a problem. However, developments in the last few years have facilitated approximate and in some cases full MLE of relatively complicated models, and there has been a corresponding interest in applications in economics and finance. At the same time, many theoretical issues are still outstanding. The design of appropriate test statistics to distinguish between I(0), I(d), and I(1) behavior is still at an early stage, and it is likely that this will be an intensively worked area in the next few years. The desire to have a semiparametric estimate of the order of fractional integration has produced a number of estimators in both the time and frequency domains. Clearly such estimators are conceptually attractive since they focus on the key parameter of interest and would ideally allow short memory effects to be neglected. However, theoretical and simulation work has generally been disappointing on the performance of these estimators; even in cases when the low-order or high-frequency dynamics are white noise.

While MLE has become straightforward computationally, the identifiability of high-order ARFIMA models often appears problematic judging from the available empirical studies and simulation evidence. In some cases the estimated value of d appears sensitive to the parametrization of the high-frequency components of the series, and in other cases the confidence interval on the estimated d parameter may include the unit root. The implication of this may be to force us to acknowledge our level of ignorance on the long-run, low-frequency properties of the data.

One great attraction of fractionally integrated processes is to allow substantially more flexibility than the extreme assumption of a unit root and its corresponding implication of the complete persistence of a shock. The slow decay of shocks implied by I(d) processes and the very slow but eventual adjustment to equilibrium is an attractive feature of the process. For example, if we are suitably skeptical about the multitude of evidence on the possible existence of a unit root in the real exchange rate, then the fractional analysis of Diebold, Husted, and Rush (1991) re-establishes Purchasing Power Parity as a meaningful long-run concept.

While many of the inferential issues in dealing with the basic ARFIMA model are now quite well understood, there are many other processes that generate long-run behavior and may be more flexible and useful than the ARFIMA class. The identification and estimation of more general long memory models, such as the Gegenbauer process or the GARMA process, are current areas of research and may become more useful representations. While most attention has currently been restricted to univariate processes, the real advantage of fractional models may well be in terms of representing relationships between variables and the testing of forms of fractional cointegration. The degree of importance of this topic will probably be determined by whether such behavior is found in real data. Some asset pricing applications certainly suggest this possibility, but little formal empirical work has been done to rigorously test this proposition.

More recently a number of authors have noted the apparent long memory property of powers of absolute returns and also of the volatility process of high frequency asset returns data. This has led to the formulation of long memory time dependent conditional heteroskedastic processes such as FIGARCH and also of corresponding long memory stochastic volatility processes. The early applied work with this approach has been extremely promising with the long memory volatility process appearing distinctly superior to other parameterizations. The feature of long memory in the conditional variance appears related to the presence of long memory in the mean of interest rate differentials and forward premia and may offer potentially important insights into the pricing of risk.

References

Adelman, I., 1965, Long cylces: Fact or artefact?, American Economic Review 55, 444–463.

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

Agiakloglou, C. and P. Newbold, 1994, Lagrange multiplier tests for fractional difference, Journal of Time Series Analysis 15, 253–262.

Agiakloglou, C., P. Newbold, and M. Wohar, 1992, Bias in an estimator of the fractional difference parameter, Journal of Time Series Analysis 14, 235–246.

Anis, A.A. and E.H. Lloyd, 1976, The expected value of the adjusted rescaled Hurst range of independent normal summands, Biometrika 63, 111–116.

Avram, F. and M.S. Taqqu, 1987, Noncentral limit theorems and Appell polynomials, Annals of Probability 15, 767–775.

Aydogan, K. and G.G. Booth, 1988, Are there long cycles in common stock returns?, Southern Economic Journal 55, 141–149.

Backus, D.K. and S.E. Zin, 1993, Long-memory inflation uncertainty: Evidence from the term structure of interest rates, Journal of Money, Credit and Banking, 681–700.

Baillie, R.T., 1989, Tests of rational expectations and market efficiency, Econometric Reviews 8, 151–186.

Baillie, R.T. and T. Bollerslev, 1989, Common stochastic trends in a system of exchange rates, Journal of Finance 44, 167–181.

Baillie, R.T. and T. Bollerslev, 1992, Prediction in dynamic models with time-dependent conditional variances, Journal of Econometrics 52, 91–113.

Baillie, R.T. and T. Bollerslev, 1994a, Cointegration, fractional cointegration and exchange rate dynamics, Journal of Finance 49, 737–745.

Baillie, R.T. and T. Bollerslev, 1994b, Long memory in the forward premium, Journal of International Money and Finance 13, 565–571.

Baillie, R.T. and R.A. Pecchenino, 1991, The search for equilibrium relationships in international finance, Journal of International Money and Finance 10, 582–593.

Baillie, R.T., T. Bollerslev, and H.-O. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, forthcoming.

Baillie, R.T., C.-F. Chung, and M.A. Tieslau, 1995, Analyzing inflation by the fractionally integrated ARFIMA-GARCH model, Journal of Applied Econometrics, forthcoming.

Beran, J.A., 1989, A test of location for data with slowly decaying serial correlations, Biometrika 76, 261–269.

Beran, J.A., 1992a, A goodness of fit test for time series with long range dependence, Journal of the Royal Statistical Society B 54, 749–760.

Beran, J.A., 1992b, Statistical methods for data with long-range dependence, Statistical Science 7, 404–427.

Beran, J.A. and N. Terrin, 1994, Estimation of the long-memory parameter, based on a multivariate central limit theorem, Journal of Time Series Analysis 15, 269–278.

Beveridge, 1925, Weather and harvest cycles, Economic Journal 31, 429–452.

Blough, S.R., 1992, The relationship between power and level for generic unit root tests in finite samples, Journal of Applied Econometrics 7, 295–308.

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

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Bollerslev, T. and H.-O. Mikkelsen, 1996, Modeling and pricing long-memory in stock market volatility, Journal of Econometrics, this issue.

Bollerslev, T., R.Y. Chou, and K.F. Kroner, 1992, ARCH modeling in finance, Journal of Econometrics 52, 5–59.

Booth, G.G., F.R. Kaen, and P.E. Koveos, 1982, R/S analysis of foreign exchange rates under two international money regimes, Journal of Monetary Economics 10, 407–415.

Breidt, F.J., N. Crato, and P.J.F. de Lima, 1993, Modeling long-memory stochastic volatility, Working paper (Johns Hopkins University, Baltimore, MD).

Brockwell, P.J. and R. Davis, 1987, Time series: Theory and methods (Springer-Verlag, New York, NY).

Campbell, J.Y. and N.G. Mankiw, 1987, Are output fluctuations transitory?, Quarterly Journal of Economics 102, 857–880.

Campbell, J.Y. and R. Shiller, 1987, Cointegration and tests of present value models, Journal of Political Economy 95, 1062–1088.

Carlin, J.B. and P. Dempster, 1989, Sensitivity analysis of seasonal adjustments: Empirical case studies, Journal of the American Statistical Association 84, 6–32.

Carlin, J.B., P. Dempster, and A.B. Jonas, 1985, On methods and models for Bayesian time series analysis, Journal of Econometrics 30, 67–90.

Chen, G., B. Abraham, and S. Peiris, 1994, Lag window estimation of the degree of differencing in fractionally integrated time series models, Journal of Time Series Analysis 15, 473–487.

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

Cheung, Y.-W., 1993b, Tests for fractional integration: A Monte Carlo investigation, Journal of Time Series Analysis 14, 331–345.

Cheung, Y.-W. and F.X. Diebold, 1994, On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean, Journal of Econometrics 62, 301–316.

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.

Choi, S. and M.E. Wohar, 1992, The performance of the GPH estimator of the fractional difference parameter, Review of Quantitative Finance and Accounting 2, 409–417.

Chung, C.-F., 1994a, A note on calculating the autocovariances of fractionally integrated ARMA models, Economics Letters 45, 293–297.

Chung, C.-F., 1994b, A generalized fractionally integrated ARMA process, Working paper (Michigan State University, East Lansing, MI).

Chung, C.-F., 1996, On estimating a generalized long memory model, Journal of Econometrics, this issue.

Chung, C.-F. and R.T. Baillie, 1993, Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models, Empirical Economics 18, 791–806.

Crato, N. and P.J.F. de Lima, 1994, Long-range dependence in the conditional variance of stock returns, Economics Letters 45, 281–285.

Crato, N. and P. Rothman, 1994a, Fractional integration analysis of long-run behavior for US macroeconomic time series, Economics Letters 45, 287–291.

Crato, N. and P. Rothman, 1994b, A reappraisal of parity reversion for UK real exchange rates, Applied Economics Letters 1, 139–141.

Dacorogna, M.M., U.A. Muller, R.J. Nagler, R.B. Olsen, and O.V. Pictet, 1993, A geographical model for the daily and weekly seasonal volatility in the foreign exchange market, Journal of International Money and Finance 12, 413–438.

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.

Davies, R.B. and D.S. Harte, 1987, Tests for Hurst effect, Biometrika 74, 95–102.

Davydov, Y.A., 1970, The invariance principle for stationary processes, Theory of Probability and its Applications 15, 487–489.

de Haan, L., 1990, Fighting the arch-enemy with mathematics, Statistica Neerlandica 44, 45–68.

Diba, B.T. and H.T. Grossman, 1988, Explosive rational bubbles in stock prices?, American Economic Review 78, 520–530.

Diebold, F.X., 1989, Random walks versus fractional integration: Power comparisons of scalar and joint tests of the variance-time function, in: B. Raj, ed., Advances in econometrics (Kluwer Academic Publishers, Needham, MA) 29–45.

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, 1991a, Is consumption too smooth? Long memory and the Deaton paradox, Review of Economics and Statistics 73, 1–9.

Diebold, F.X. and G.D. Rudebusch, 1991b, On the power of Dickey–Fuller tests against fractional alternatives, Economics Letters 35, 155–160.

Diebold, F.X., J. Gardeazabal, and K. Yilmaz, 1994, On cointegration and exchange rate dynamics, Journal of Finance 49, 727–735.

Diebold, F.X., S. Husted, and M. Rush, 1991, Real exchange rates under the gold standard, Journal of Political Economy 99, 1252–1271.

Ding, Z., C.W.J. Granger, and R.F. Engle, 1993, A long memory property of stock returns and a new model, Journal of Empirical Finance 1, 83–106.

Engel, C., 1991, On the foreign exchange risk premium in a general equilibrium model, Journal of International Economics 32, 305–319.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica 50, 987–1008.

Evans, M.D.D. and K.K. Lewis, 1993, Do long term swings in the dollar affect estimates of the risk premia?, Working paper (University of Pennsylvania, Philadelphia, PA).

Fama, E.F., 1984, Spot and forward exchange rates, Journal of Monetary Economics 14, 319–338.

Faust, J., 1994, Near observational equivalence and unit root processes: Formal concepts and implications, Working paper (Board of Governors of the Federal Reserve System, Washington, DC).

Feller, W., 1951, The asymptotic distribution of the range of sums of independent random variables, Annals of Mathematical Statistics 22, 427–432.

Fox, R. and M.S. Taqqu, 1985, Noncentral limit theorems for quadratic forms in random variables having long range dependence, Annals of Probability 13, 428–446.

Fox, R. and M.S. Taqqu, 1986, Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, Annals of Statistics14, 517–532.

Fox, R. and M.S. Taqqu, 1987, Central limit theorems for quadratic forms in random variables having long-range dependence, Probability Theory and Related Fields 74, 213–240.

Galbraith, R.F. and J.F. Galbraith, 1974, On the inverse of some patterned matrices arising in the theory of stationary time series, Journal of Applied Probability 11, 63–71.

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.

Giraitis, L. and D. Surgailis, 1990, A central limit theorem for quadratic form in strongly dependent linear variables and its application to Whittle's estimate, Probability Theory and Related Fields 86, 87–104.

Gradszteyn, I.S. and I.M. Ryzhnik, 1980, Tables of integrals, series and products, 4th ed. (Academic Press, New York, NY).

Granger, C.W.J., 1966, The typical spectral shape of an economic variable, Econometrica 34, 150–161.

Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14, 227–238.

Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, Journal of Econometrics 16, 121–130.

Granger, C.W.J., 1983, Cointegrated variables and error correction models, Unpublished manuscript (University of California, San Diego, CA).

Granger, C.W.J. and A. Andersen, 1978, On the invertibility of time series models, Stochastic Process Applications 8, 87–92.

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.

Gray, H.L., N.-F. Zhang, and W.A. Woodward, 1989, On generalized fractional processes, Journal of Time Series Analysis 10, 233–257.

Greene, M. and B. Fielitz, 1977, Long-term dependence in common stock returns, Journal of Financial Economics 4, 339–349.

Hakkio, C.S. and M. Rush, 1991, Cointegration: How short is the long run?, Journal of International Money and Finance 10, 571–581.

Harvey, A.C., 1993, Long memory in stochastic volatility, Working paper (London School of Economics, London).

Haslett, J. and A.E. Raftery, 1989, Space–time modelling with long-memory dependence: Assessing Ireland's wind power resource, Journal of the Royal Statistical Society C 38, 1–50.

Hassler, U., 1993, Regression of spectral estimators with fractionally integrated time series, Journal of Time Series Analysis 14, 369–380.

Hassler, U., 1994, (Mis)specification of long memory in seasonal time series, Journal of Time Series Analysis 15, 19–30.

Hassler, U. and J. Wolters, 1994, On the power of unit roots against fractionally integrated alternatives, Economics Letters 45, 1–5.

Hassler, U. and J. Wolters, 1995, Long memory in inflation rates: International evidence, Journal of Business and Economic Statistics 13, 37–45.

Haubrich, J.G., 1992, Consumption and fractional differencing: Old and new anomalies, Working paper (Wharton School, University of Pennsylvania, Philadelphia, PA).
Haubrich, J.G. and A.W. Lo, 1993, The sources and nature of long-term memory in the business cycle, Working paper (Federal Reserve Bank of Cleveland, Cleveland, OH).
Hauser, M.A., 1994, Semiparametric and nonparametric testing for long memory, Working paper (University of Vienna, Vienna).
Helson, J. and Y. Sarason, 1967, Past and future, Mathematica Scandanavia 21, 5–16.
Hipel, K.W. and A.I. McLeod, 1978a, Preservation of the rescaled adjusted range, 3: Fractional Gaussian noise algorithms, Water Resources Research 14, 517–518.
Hipel, K.W. and A.I. McLeod, 1978b, Preservation of the rescaled adjusted range, 2: Simulation studies using Box–Jenkins models, Water Resources Research 14, 509–516.
Hols, M.C.A.B. and C.G. de Vries, 1991, The limiting distribution of extremal exchange rate returns, Journal of Applied Econometrics 6, 287–302.
Hosking, J.R.M., 1981, Fractional differencing, Biometrika 68, 165–176.
Hosking, J.R.M., 1984, Modeling persistence in hydrological time series using fractional differencing, Water Resources Research 20, 1898–1908.
Hurst, H.E., 1951, Long-term storage capacity of reserviors, Transactions of the American Society of Civil Engineers 116, 770–799.
Hurst, H.E., 1956, Methods of using long term storage in reservoirs, Proceedings of the Institute of Civil Engineers 1, 519–543.
Hurst, H.E., 1957, A suggested statistical model of some time series that occur in nature, Nature 180, 494.
Hurvich, C.M. and K.I. Beltrao, 1994, Automatic semiparametric estimation of the long memory parameter of a long memory time series, Journal of Time Series Analysis 15, 285–302.
Hurvich, C.M. and B.K. Ray, 1995, Estimation of the memory parameter for non-stationary or non-invertible fractionally integrated processes, Journal of Time Series Analysis 16, 17–41.
Janacek, G.J., 1982, Determining the degree of differencing for time series via the log spectrum, Journal of Time Series Analysis 3, 177–183.
Kaen, F.R. and R.E. Rosenman, 1986, Predictable behavior in financial markets: Some evidence in support of Heinber's hypothesis, American Economic Review 76, 212–220.
Kennedy, D., 1976, The distribution of the maximum Brownian excursion, Journal of Applied Probability 13, 371–376.
Kim, Y., 1990, Purchasing power parity in the long run: A cointegration approach, Journal of Money, Credit and Banking 22, 491–503.
King, M.L. and G. Hillier, 1985, Locally best invariant tests of the error covariance matrix of the linear regression model, Journal of the Royal Statistical Society B 47, 98–102.
Koedijk, K.G., M.M.A. Schafgans, and C.G. de Vries, 1990, The tail index of exchange rate returns, Journal of International Economics 29, 93–108.
Kolmogorov, A.N., 1940, Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum, Comptes Rendues (Doklady) URSS NS 26, 115–118.
Kunsch, H., 1986, Discrimination between monotonic trends and long-range dependence, Journal of Applied Probability 23, 1025–1030.
Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159–178.
Lawrance, A.J. and N.T. Kottegoda, 1977, Stochastic modelling of riverflow time series, Journal of the Royal Statistical Society A 140, 1–47.
Leadbetter, M.R., G. Lindgren, and H. Rootzen, 1984, Extremes and related properties of random sequences and processes (Springer-Verlag, New York, NY).
Lee, D. and P. Schmidt, 1996, On the power of the KPSS test of stationarity against fractionally integrated alternatives, Journal of Econometrics, this issue.

Li, W.K. and A.I. McLeod, 1986, Fractional time series modeling, Biometrika 73, 217–221.

Lin, J.-Y., 1991, Generalized integrated process and the aggregation of dynamic time series, Academia Economic Papers 19, 207–226.

Lo, A.W., 1991, Long term memory in stock market prices, Econometrica 59, 1279–1313.

Lobato, I. and P.M. Robinson, 1996, Averaged periodogram estimation of long memory, Journal of Econometrics, this issue.

MacNeill, I., 1978, Properties of sequences of partial sums of polynomial regression residuals with application to tests of change of regression at unknown times, Annals of Statistics 6, 422–433.

McLeod, A.I. and K.W. Hipel, 1978, Preservation of the rescaled adjusted range, 1: A reassessment of the Hurst phenomenon, Water Resources Research 14, 491–508.

Mandelbrot, B.B., 1971, When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models, Review of Economics and Statistics 53, 225–236.

Mandelbrot, B.B., 1972, Statistical methodology for non periodic cycles: From the covariance to R/S analysis, Annals of Economic and Social Measurement 1, 259–290.

Mandelbrot, B.B., 1975, A fast fractional Gaussian noise generator, Water Resources Research 7, 543–553.

Mandelbrot, B.B. and M. Taqqu, 1979, Robust R/S analysis of long run serial correlation, Bulletin of International Statistical Institute 48, book 2, 59–104.

Mandelbrot, B.B. and J.W. Van Ness, 1968, Fractional Brownian motions, fractional Brownian noises and applications, SIAM Review 10, 422–437.

Mandelbrot, B.B. and J. Wallis, 1968, Noah, Joseph and operational hydrology, Water Resources Research 4, 909–918.

Mandelbrot, B.B. and J. Wallis, 1969a, Computer experiments with fractional Gaussian noises, Parts 1,2,3, Water Resources Research 5, 228–267.

Mandelbrot, B.B. and J. Wallis, 1969b, Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence, Water Resources Research 5, 967–988.

Meese, R.A. and K.L. Singleton, 1981, On unit roots and the empirical modeling of exchange rates, Journal of Finance 37, 1029–1035.

Moehring, R., 1990, Parameter estimation in Gaussian intermediate memory time series, Working paper (Institut für Mathematische Stochastik, University of Hamburg, Hamburg).

Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347–370.

Nelson, C.R. and C.I. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, Journal of Monetary Economics 10, 139–162.

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

Newbold, P. and C. Agiakloglou, 1993, Bias in the sample autocorrelations of fractional white noise, Biometrika 80, 698–702.

Noakes, D.J., K.W. Hipel, A.I. McLeod, C. Jimenez, and S. Yakowitz, 1988, Forecasting annual geophysical time series, International Journal of Forecasting 4, 103–115.

Odaki, M., 1993, On the invertibility of fractionally differenced ARIMA processes, Biometrika 80, 703–709.

Peiris, M.S., 1987, A note on the predictors of differenced sequences, Australian Journal of Statistics 29, 42–48.

Peiris, M.S. and B.J.C. Perera, 1988, On prediction with fractionally differenced ARMA models, Journal of Time Series Analysis 9, 215–220.

Phillips, P.C.B., 1987, Time series regression with a unit root, Econometrica 55, 277–301.

Porter-Hudak, S., 1990, An application of the seasonally fractionally differenced model to the monetary aggregates, Journal of the American Statistical Association 85, 338–344.

Ray, B.K., 1993a, Modeling long-memory processes for optimal long-range prediction, Journal of Time Series Analysis 14, 511–525.

Ray, B.K., 1993b, Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model, International Journal of Forecasting 9, 255-269.
Reisen, V.A., 1994, Estimation of the fractional difference parameter in the ARFIMA(p,d,q) model using the smoothed periodogram, Journal of Time Series Analysis 15, 335-351.
Resnick, 1987, Extreme values, regular variation and point processes (Springer-Verlag, New York, NY).
Robinson, P.M., 1978, Statistical inference for a random coefficient autoregressive model, Scandanavian Journal of Statistics 5, 163-168.
Robinson, P.M., 1990, Time series with strong dependence, Advances in econometrics, 6th world congress (Cambridge University Press, Cambridge).
Robinson, P.M., 1991, Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, Journal of Econometrics 47, 67-84.
Robinson, P.M., 1992, Semiparametric analysis of long-memory time series, Annals of Statistics 22, 515-539.
Rosenblatt, M., 1956, A central limit theorem and a strong mixing condition, Proceedings of the National Academy of Sciences 42, 43-47.
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.
Schmidt, P. and P.C.B. Phillips, 1992, LM tests for a unit root in the presence of deterministic trend, Oxford Bulletin of Economics and Statistics 54, 257-287.
Seater, J.J., 1993, World temperature-trend uncertainties and their implications for economic policy, Journal of Business and Economic Statistics 11, 265-277.
Sephton, P.S. and H.K. Larsen, 1991, Tests of exchange market efficiency: Fragile evidence from cointegration tests, Journal of International Money and Finance 10, 561-570.
Shea, G.S., 1991, Uncertainty and implied variance bounds in long memory models of the interest rate term structure, Empirical Economics 16, 287-312.
Shiller, R.J. and P. Perron, 1985, Testing the random walk hypothesis: Power versus frequency of observations, Economics Letters 18, 381-386.
Siddiqui, M., 1976, The asymptotic distribution of the range and other functions of partial sums of stationary processes, Water Resources Research 12, 1271-1276.
Sinai, Y.G., 1976, Self-similar probability distributions, Theory of Probability and its Applications 21, 64-80.
Smith, F.B. Sowell, and S.E. Zin, 1993, Fractional integration with drift: Estimation in small samples, Working paper (Carnegie Mellon University, Pittsburgh, PA).
Sowell, F.B., 1986, Fractionally integrated vector time series, Ph.D. dissertation (Duke University, Durham, NC).
Sowell, F.B., 1990, The fractional unit root distribution, Econometrica 58, 495-505.
Sowell, F.B., 1992a, Maximum likelihood estimation of stationary univarite fractionally integrated time series models, Journal of Econometrics 53, 165-188.
Sowell, F.B., 1992b, Modeling long run behavior with the fractional ARIMA model, Journal of Monetary Economics 29, 277-302.
Steigerwald, D.G., 1994, Purchasing power parity, unit roots and dynamic structure, Journal of Empirical Finance, forthcoming.
Taqqu, M.S., 1975, Weak convergence to fractional Brownian motion and to the Rosenblatt process, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31, 287-302.
Taqqu, M.S., 1977, Law of the iterated logarithms for sums of non-linear functions of Gaussian variables that exhibit a long range dependence, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 40, 203-238.
Taylor, S., 1986, Modeling financial time series (Wiley, Chichester).
Tieslau, M.A., 1992, Strongly dependent economic time series: Theory and applications, Ph.D. dissertation (Michigan State University, East Lansing, MI).

Tieslau, M.A., P. Schmidt, and R.T. Baillie, 1995, A minimum-distance estimator for long memory processes, Journal of Econometrics, forthcoming.

Tschernig, R., 1992, Wechselkurse, Unsicherheit und Long Memory, Ph.D. dissertation (University of Munich, Munich).

Whittle, P., 1951, Hypothesis testing in time series analysis (Almquist and Wiksells, Uppsala).

Whittle, P., 1956, Variation of yield variance with plot size, Biometrika 43, 337–343.

Wu, P., 1992, Testing fractionally integrated time series, Working paper (Victoria University, Wellington).

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.

Yajima, Y., 1989, A central limit theorem of Fourier transforms of strongly dependent stationary processes, Journal of Time Series Analysis 10, 373–383.