# THE ESTIMATION AND APPLICATION OF LONG MEMORY TIME SERIES MODELS

BY JOHN GEWEKE AND SUSAN PORTER-HUDAK[1]

*Department of Economics, Duke University*

**Abstract.** The definitions of fractional Gaussian noise and integrated (or fractionally differenced) series are generalized, and it is shown that the two concepts are equivalent. A new estimator of the long memory parameter in these models is proposed, based on the simple linear regression of the log periodogram on a deterministic regressor. The estimator is the ordinary least squares estimator of the slope parameter in this regression, formed using only the lowest frequency ordinates of the log periodogram. Its asymptotic distribution is derived, from which it is evident that the conventional interpretation of these least squares statistics is justified in large samples. Using synthetic data the asymptotic theory proves to be reliable in samples of 50 observations or more. For three postwar monthly economic time series, the estimated integrated series model provides more reliable out-of-sample forecasts than do more conventional procedures.

**Keywords.** Fractional differencing, Long-memory, Integrated models

## 1. INTRODUCTION

In the most widely applied models for stationary time series the spectral density function is bounded at the frequency $\lambda = 0$ and the autocorrelation function decays exponentially. This is true in stable autoregressive moving average models, and it is characteristic of estimates obtained by nonparametric spectral methods, for example. Yet these properties do not appear widely characteristic of many time series (Hurst, 1951; Granger and Joyeux, 1980). The failure of these models to represent the spectral density at low frequencies adequately suggests that many-step-ahead forecasts obtained using these models could be inferior to those produced by models that permit unbounded spectral densities at $\lambda = 0$ and autocorrelation functions that do not decay exponentially. Limited empirical evidence to this effect has been reported by Granger and Joyeux, and is supported by somewhat more extensive investigations reported by Porter-Hudak (1982).

Attention has recently been given to two single-parameter models in which the spectral density function is proportional to $\lambda^{-r}$, $1 < r < 2$, for $\lambda$ near 0, and the asymptotic decay of the autocorrelation function is proportional to $\tau^{r-1}$. Because the spectral density function is unbounded at $\lambda = 0$—equivalently, the autocorrelation function is not summable—these are *long memory models* (defined by McLeod and Hipel, 1978).

The earlier model was introduced by Mandelbrot and Van Ness (1968) and Mandelbrot (1971) to formalize Hurst's empirical findings using cumulative river flow data. Let $Y_t = \int_{-\infty}^{t} (t-s)^{H-1/2}\, dB(s)$, where $B(s)$ is Brownian motion and $H \in (0, 1)$. Then $X_t = Y_t - Y_{t-1}$ is a *simple fractional Gaussian noise.* (We have

0143-5782/83/04 0221-18 $02.50/0
JOURNAL OF TIME SERIES ANALYSIS Vol. 4, No. 4
© 1983 V. Geweke, S. Porter-Hudak

added the word 'simple' to the definition given in the literature, to emphasize the difference between that model and the extension of it introduced here.) Jonas (1981) has shown that its spectral density is

$$f_1(\lambda; H) = \sigma^2(2\pi)^{-2H-2}\Gamma(2H+1)\sin(\pi H)4\sin^2(\lambda/2)\sum_{n=-\infty}^{\infty}|n+(\lambda/2\pi)|^{-2H-1}.$$

(We use the convention that the spectral density of a time series $\{X_t\}$ is $f(\lambda) = \sum_{s=-\infty}^{\infty} R_x(s)\exp(-i\lambda s)$, where $R_x(s)$ is the autocovariance function of $\{X_t\}$.) A little manipulation of this expression shows that

$$\lim_{\lambda\to 0}\lambda^{2H-1}f_1(\lambda; H) = (2\sigma^2/\pi)\Gamma(2H+1)\sin(\pi H).$$

The corresponding autocorrelation function is

$$\rho_1(\tau; H) = .5(|\tau-1|^{2H} - 2|\tau|^{2H} + |\tau+1|^{2H}).$$

A Taylor series expansion in $(\tau-1)/\tau$ and $(\tau+1)/\tau$ shows

$$\lim_{\tau\to\infty}\tau^{2-2H}\rho_1(\tau; H) = H(2H-1),$$

and from the Schwarz inequality $\rho_1(\tau; H) \gtrless 0$ for $\tau \neq 0$ as $H \gtrless .5$. In empirical work interest has centered on the cases $.5 < H < 1$.

The fractional Gaussian noise model is designed to account for the long term behaviour of the time series in question. Realistically, it seems unlikely that all the second moment properties of the series would be well described by any single-parameter function, but shorter term behaviour could be modelled by more conventional means. Toward this end, $\{X_t\}$ will be said to be a *general fractional Gaussian noise* if its spectral density is of the form $f_1(\lambda; H)f_u(\lambda)$ where $f_u(\lambda)$ is a positive continuous function bounded above and away from zero on the interval $[-\pi, \pi]$. The spectral density $f_u(\lambda)$ therefore is that of a short memory time series (McLeod and Hippel) with autoregressive representation (Rozanov, 1967, pp. 77–78). This class of models is broad: it includes, for example, time series arising from $\phi(B)X_t = \theta(B)X_t'$, where $\phi(B)$ and $\theta(B)$ are invertible polynomials of finite order in the lag operator $B$ and $X_t'$ is a simple fractional Gaussian noise.

The second long memory model was proposed independently by Granger and Joyeux (1980) and Hosking (1981). It can be motivated by the observation that some time series appear to have unbounded spectral densities at the frequency $\lambda = 0$, but the spectral densities of first differences of the same series appear to vanish at $\lambda = 0$. This suggests the model $(1-B)^d X_t = \varepsilon_t$, where $d \in (-.5, .5)$ and $\varepsilon_t$ is serially uncorrelated. By this it is meant that the spectral density of $\{X_t\}$ is

$$f_2(\lambda; d) = (\sigma^2/2\pi)|1-e^{-i\lambda}|^{-2d} = (\sigma^2/2\pi)\{4\sin^2(\lambda/2)\}^{-d}.$$

A series with the spectral density $f_2(\lambda; d)$ will be called a *simple integrated series.* Clearly $\lim_{\lambda\to 0}\lambda^{2d}f_2(\lambda; d) = (\sigma^2/2\pi)$. The corresponding autocorrelation func-

tion (for $d \neq 0$) is

$$\rho_2(\tau; d) = \Gamma(1-d)\Gamma(\tau+d)/\{\Gamma(d)\Gamma(\tau+1-d)\},$$

and $\lim_{\tau\to\infty} \tau^{1-2d}\rho_2(\tau; d) = \Gamma(1-d)/\Gamma(d) = \pi/\sin(\pi d)$ (Granger and Joyeux, 1980, p. 17). A simple integrated series has the autoregressive representation

$$\sum_{j=0}^{\infty} a_2(j, d)X_{t-j} = \varepsilon_t, \qquad a_2(j; d) = \Gamma(j-d)/\{\Gamma(-d)\Gamma(j+1)\},$$

and moving average representation

$$X_t = \sum_{j=0}^{\infty} b_2(j; d)\varepsilon_{t-j}, \qquad b_2(j; d) = \Gamma(j+d)/\{\Gamma(j+1)\Gamma(d)\}$$

(Hosking, 1981, p. 167; Granger and Joyeux, 1980, equation (6) is in error). Since $\lim_{j\to\infty} j^{(1+d)}a_2(j; d) = 1/\Gamma(-d)$ and $\lim_{j\to\infty} b_2(j; d) = 1/\Gamma(d)$, $a_2(j; d)$ and $b_2(j; d)$ are each square summable; but $\sum_{j=0}^{\infty}|a_2(j; d)| > \infty$ if and only if $d > 0$, while $\sum_{j=0}^{\infty}|b_2(j; d)| < \infty$ if and only if $d < 0$.

It seems desirable to extend the integrated series model in the same way as the fractional Gaussian noise model. Hence $\{X_t\}$ will be called a *general integrated series* if its spectral density is of the form $f_2(\lambda; d)f_u(\lambda)$, where $f_u(\lambda)$ has the same characteristics as before. Whenever the process $\{\varepsilon_t\}$ is i.i.d. as well as being serially uncorrelated, we shall add the adjective *linear* to describe the process.

There are obvious similarities in simple fractional Gaussian noise and simple integrated series, to which Granger and Joyeux (1980), Hosking (1981), and Jonas (1981) have referred obliquely. In section 2 it is shown that the spectral density function of a general fractional Gaussian noise with parameter $H$ is that of a general integrated series with parameter $H-\frac{1}{2}$, and vice versa. Which model one uses therefore depends on practical considerations, to which the rest of the paper is addressed. In section 3 a consistent, computationally efficient estimator $\hat{d}$ of the parameter $d$ for general integrated series is introduced, and its asymptotic distribution is derived. (By virtue of the results of section 2, $\hat{H} = \hat{d}+\frac{1}{2}$ provides a consistent estimator of $H$, and $\hat{H}-H$ and $\hat{d}-d$ have the same asymptotic distribution.) Results with synthetic time series are reported in section 4, and with actual time series in section 5. On the basis of these results it can be concluded that the estimator $\hat{d}$ is at least as reliable in finite sample as any estimator of $d$ and $H$ suggested to date, and is more attractive computationally.

## 2. THE EQUIVALENCE OF GENERAL FRACTIONAL GAUSSIAN NOISE AND GENERAL INTEGRATED SERIES

We provide the following characterization of the relationship of the two models.

THEOREM 1. *$\{X_t\}$ is a general integrated series with parameter $d$ $(-\frac{1}{2} < d < \frac{1}{2})$ if, and only if, it is also a general fractional Gaussian noise with parameter $H = d+\frac{1}{2}$.*

PROOF. Let $d=(L-2)/2$ and $H=(L-1)/2$, $1<L<3$.

$$f_1\{\lambda;(L-1)/2\}/f_2\{\lambda;(L-2)/2\}$$

$$=\pi^{-1}\Gamma(L)\sin\{\pi(L-1)/2\}|\sin(\lambda/2)|^L\sum_{n=-\infty}^{\infty}|n+(\lambda/2\pi)|^{-L}$$

$$=\pi(L)\sin\{\pi(L-1)/2\}|\sin(\lambda/2)/(\lambda/2)|^L\sum_{n=-\infty}^{\infty}|(\lambda/2\pi)/\{n+(\lambda/2\pi)\}|^L.$$

It suffices to show that this ratio is continuous in $\lambda$ and bounded above and below by positive numbers on $[-\pi,\pi]$. The function $|\sin(\lambda/2)/(\lambda/2)|^L$ has these characteristics. Furthermore

$$1<\sum_{n=-\infty}^{\infty}|(\lambda/2\pi)/\{n+(\lambda/2\pi)\}|^L$$

$$=1+\sum_{n=1}^{\infty}|(\lambda/2\pi)/\{n+(\lambda/2\pi)\}|^L+\sum_{n=1}^{\infty}|(\lambda/2\pi)/\{n-(\lambda/2\pi)\}|^L$$

$$<1+2|(\lambda/2\pi)/\{1+(\lambda/2\pi)\}|^L\zeta(L),$$

where $\zeta(L)=\sum_{n=1}^{\infty}n^{-L}$ is Riemann's zeta function; $\zeta(L)<\infty$ for $L>1$. The function $\sum_{n=-\infty}^{\infty}|(\lambda/2\pi)/\{n+(\lambda/2\pi)\}|^L$ is therefore continuous and bounded above and below by positive numbers on $[-\pi,\pi]$ also.

The proof shows that $f_1(\lambda;(L-1)/2)/f_2(\lambda;(L-2)/2)$ is bounded uniformly in $\lambda$. It is not bounded uniformly in $L$, and its behaviour as $L\to1$ and $L\to3$ indicates potential complications in applied work. To examine the case $L\to1$, write

$$f_1\{\lambda;(L-1)/2\}/f_2\{\lambda;(L-2)/2\}$$

$$=\sin(\lambda/2)\pi^{-1}\Gamma(L)\frac{\sin\{\pi(L-1)/2\}}{\sin(\pi L)}\sin(\pi L)\sum_{n=-\infty}^{\infty}|n+(\lambda/2\pi)|^{-L}.$$

Substitute $\sin(\pi L)=\pi/\{\Gamma(L)\Gamma(1-L)\}$ (Gradshteyn and Ryzhik, 1980, 8.334.3) and $\sum_{n=-\infty}^{\infty}|n+(\lambda/2\pi)|^{-L}=\zeta(L,\lambda/2\pi)+\zeta(L,-\lambda/2\pi)-(-|\lambda|/2\pi)^{-L}$ (where $\zeta(L,q)=\sum_{n=0}^{\infty}(n+q)^{-L}$ is Riemann's general zeta function) to obtain

$$\sin(\lambda/2)\pi^{-L+1}\Gamma(L)\frac{\sin\{\pi(L-1)/2\}}{\sin(\pi L)\Gamma(L)\Gamma(1-L)}$$

$$\times\{\zeta(L,\lambda/2\pi)+\zeta(L,-\lambda/2\pi)-(-|\lambda|/2\pi)^{-L}\}.$$

Since $\lim_{L\to1}\zeta(L,q)/\Gamma(1-L)=-1$ (Gradshteyn and Ryzhik, 1980, 9.533.1) the limiting value of this expression as $L\to1$ is $\sin(\lambda/2)$. Hence if a general integrated series were to be described as a general fractional Gaussian noise, then as $d\to-\frac{1}{2}$ a noninvertible moving average component would be required.

As $L\to3$ the ratio approaches zero, but the approach is uniform in $\lambda$: $\lim_{L\to3}f_1\{\lambda;(L-1)/2\}/[f_2(\lambda;(L-2)/2)\sin\{\pi(L-1)/2\}]$ is continuous and is bounded above and below by positive constants. This limiting case seems to pose no practical problems.

## 3. A SIMPLE ESTIMATION PROCEDURE FOR GENERAL INTEGRATED SERIES

Consider the problem of estimating the parameter $d$ in the general integrated series model. Suppose $(1-B)^d X_t = u_t$, where $u_t$ is a stationary linear process with spectral density function $f_u(\lambda)$ which is finite, bounded away from zero and continuous on the interval $[-\pi, \pi]$. The spectral density function of $\{X_t\}$ is $f(\lambda) = (\sigma^2/2\pi)\{4\sin^2(\lambda)\}^{-d} f_u(\lambda)$, and

$$\ln\{f(\lambda)\} = \ln\{\sigma^2 f_u(0)/2\pi\} - d\ln\{4\sin^2(\lambda/2)\} + \ln\{f_u(\lambda)/f_u(0)\}. \quad (1)$$

Suppose that a sample of $\{X_t\}$ of size $T$ is available. Let $\lambda_{j,T} = 2\pi j/T$ $(j = 0, \ldots, T-1)$ denote the harmonic ordinates, and $I(\lambda_{j,T})$ denote the periodogram at these ordinates. Evaluate (1) at $\lambda_{j,T}$ and rearrange to obtain

$$\ln\{I(\lambda_{j,T})\} = \ln\{\sigma^2 f_u(0)/2\pi\} - d\ln\{4\sin^2(\lambda_{j,T}/2)\} + \ln\{f_u(\lambda_{j,T})/f_u(0)\} + \ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}. \quad (2)$$

The proposed estimate of $d$ is motivated by the formal similarity of (2) and a simple linear regression equation: $\ln\{I(\lambda_{j,T}/2)\}$ is analogous to the dependent variable, $\ln\{4\sin^2(\lambda_{j,T})\}$ is the explanatory variable, $\ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}$ is the disturbance, the slope coefficient is $-d$, and the intercept term is $\ln\{\sigma^2 f_u(0)/2\pi\}$ plus the mean of $\ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}$. The term $\ln\{f_u(\lambda_{j,T})/f_u(0)\}$ becomes negligible as attention is confined to harmonic frequencies nearer to zero. The proposed estimator is the slope coefficient in the least squares regression of $\ln\{I(\lambda_{j,T})\}$ on a constant and $\ln\{4\sin^2(\lambda_{j,T}/2)\}$ in the sample $j = 1, \ldots, g(T)$; $g(T)$ will be described subsequently. It will be shown that when $d < 0$ the estimator is consistent, and the conventional interpretation of the standard error for the slope coefficient is appropriate asymptotically. We conjecture that the results remain true for $d \geq 0$, and experimental evidence to that effect will be provided subsequently.

To begin, consider (2) for $j = 1, \ldots, n$, for any arbitrarily chosen positive integer $n$. As $T$ increases, the harmonic frequencies $\lambda_{j,T}$ all approach zero and the term $\ln\{f_u(\lambda_{j,T})/f_u(0)\}$ may be ignored. When $d < 0$ the coefficients in the moving average representation of $\{X_t\}$ are absolutely summable. Hence the random variables $Z_{j,T} = \sum_{t=1}^{T} X_t \exp(-i\lambda_{j,T})/\{f(\lambda_{j,T})\}^{1/2}$ are asymptotically i.i.d. normal (Hannan, 1973, theorem 3). The terms $\ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}$ are therefore also asymptotically i.i.d., and their distribution can be derived by change of variable techniques (Porter-Hudak, 1982, Appendix C). The distribution is of the Gumbel type; the asymptotic mean of $\ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}$ is $-C$ ($C$ is Euler's constant, .57721 . . . ) and its variance is $\pi^2/6$. Consider invoking a conventional central limit theorem to obtain the asymptotic distribution of the least squares slope coefficient in (2). This requires that certain conditions on the regressors be met (to be discussed shortly) and that the disturbance term be i.i.d. with finite mean and variance. This condition will generally never be met for fixed $n$ *or* $T$, but for fixed $n$ it may be approximated to any specified degree of accuracy by choosing $T$ sufficiently large. Once the accuracy criterion is met the number of ordinates can be increased to $n+1$, and $T$ can then be further increased until

the criterion of accuracy is again met. In this way a function $g(T)$ is defined, such that if $n = g(T)$ then the least squares slope estimator $\hat{d}$ of $d$ is asymptotically normal. Clearly $g(T)$ must satisfy $\lim_{T\to\infty} g(T) = \infty$, $\lim_{T\to\infty} g(T)/T = 0$, but beyond this all that is known is that $g(T)$ exists.

Let $U(j, T) = \ln\{4\sin^2(\lambda_{j,T}/2)\}$, $\bar{U}(T, n) = n^{-1}\sum_{j=1}^{n} U(j, T)$. The asymptotic normality of $\hat{d}$ follows from the Lindberg-Levy central limit theorem (Kendall and Stuart, 1972, pp. 206–208).

$$\lim_{T\to\infty} \sum_{j=1}^{g(T)} \{U(j, T) - \bar{U}(T, g(T))\}^2 = \infty, \tag{3}$$

$$\lim_{T\to\infty} \{U(1, T)\}^2 \Big/ \left[\sum_{j=1}^{g(T)} \{U(j, T) - \bar{U}(T, g(T))\}^2\right] = 0. \tag{4}$$

(In (4), use has been made of the fact that $\sup_{j=1,\ldots,g(T)} \{U(j, T)\}^2 = \{U(1, T)\}^2$.) In showing (3) and (4) it is simpler to work with $W(j, T) = \ln(\lambda_{j,T})$ and $\bar{W}(T, n) = n^{-1}\sum_{j=1}^{n} W(j, T)$ in lieu of $U(j, T)$ and $\bar{U}(T, n)$, respectively. The substitution is justified by the fact that for any $g(T)$ that satisfies $\lim_{T\to\infty} g(T)/T = 0$, $\lim_{T\to\infty} U(j, T)/W(j, T) = 2$ $(j = 1, \ldots, g(T))$ and

$$\lim_{T\to\infty} \left\{\sum_{j=1}^{g(T)} U(j, T) \Big/ \sum_{j=1}^{g(T)} W(j, T)\right\}^2 = \lim_{T\to\infty} \sum_{j=1}^{g(T)} \{U(j, T)\}^2 \Big/ \sum_{j=1}^{g(T)} \{Z(j, T)\}^2 = 4.$$

To simplify notation let $n = g(T)$. To establish (3) observe that

$$\sum_{j=1}^{n} \{W(j, T) - \bar{W}(j, T)\}^2 = \sum_{j=1}^{n} (\ln j)^2 - n^{-1}\left(\sum_{j=1}^{n} \ln j\right)^2. \tag{5}$$

Make the substitutions

$$\sum_{j=1}^{n} (\ln j)^2 = (n+\tfrac{1}{2})(\ln n)^2 - 2n \ln n + 2n + C_n$$

where $-(\ln 2)^2/2 - 2 < C_n < -2$ (Buck, 1978, p. 252), and

$$\sum_{j=1}^{n} \ln j = (n+\tfrac{1}{2}) \ln n - n + D_n,$$

where $1 - \ln 2/2 < D_n < 1$ (Buck, 1978, p. 252). Expression (5) then reduces to

$$n - (\tfrac{1}{2} + \tfrac{1}{4}n^{-1})(\ln n)^2 + (1 - 2D_n - n^{-1}D_n) \ln n + C_n + 2D_n - n^{-1}D_n^2, \tag{6}$$

which for large $n$ is dominated by $n \to \infty$. On the other hand, $\{W(1, T)\}^2 = (\ln 2\pi - \ln T)^2$. So long as

$$\lim_{T\to\infty} (\ln T)^2/g(T) = 0, \tag{7}$$

(4) will be satisfied. This appears to be a weak requirement: e.g., it is satisfied by $g(T) = cT^{\alpha}$, $0 < \alpha < 1$.

In inference about $d$ from $\hat{d}$, use can be made of the known variance of the analogue of the disturbance term in (2): in large samples, the distribution of $\hat{d}$ is approximated by

$$N\left(d, \pi^2 \Big/ \left[6\sum_{j=1}^{g(T)} \{U(j, T) - \bar{U}(T, g(T))\}^2\right]\right).$$

A natural, somewhat simpler procedure is to assume that $\hat{d}$ is normal with standard deviation given by the usual ordinary least squares arithmetic. This assumption is justified so long as $s^2$ consistently estimates $\text{var}\{I(\lambda_{j,T})/f(\lambda_{j,T})\}$ in (2), which will be guaranteed if the least squares intercept $b_0$ converges in probability to the population intercept $\ln\{\sigma^2 f_u(0)/2\pi\} - C$. Since $b_0 = n^{-1}\sum_{j=1}^{n} \ln\{I(\lambda_{j,T})\} - \hat{d}\bar{U}(T, n)$, this will occur if $\text{plim}\,(\hat{d}-d)\bar{U}(T, n) = 0$. As argued above, $\bar{U}(T, n)$ behaves like $n^{-1}\sum_{j=1}^{n} \ln(2\pi j/T) = \ln 2\pi - \ln T + \ln n + \ln n/2n - 1 + D_n/n$ when $n$ is large, so $\lim \bar{U}(T, n)/\ln(n/T) = 1$. From (6), the asymptotic standard error of $\hat{d}$ is proportional to $n^{-1/2}$. Hence $b_0$ is consistent for the population intercept, and $\text{plim}\, s^2 = \pi^2/6$, if $\lim_{T\to\infty}\{g(T)\}^{-1/2} \ln\{g(T)/T\} = 0$. This condition is implied by (7).

These results are collected in the following theorem.

THEOREM 2. *Suppose* $\{X_t\}$ *is a general integrated linear process, with* $d < 0$. *Let* $I(\lambda_{j,T})$ *denote the periodogram of* $\{X_t\}$ *at the harmonic frequencies* $\lambda_{j,T} = \pi j/T$ *in a sample of size* $T$. *Let* $b_{1,T}$ *denote the ordinary least squares estimator of* $\beta_1$ *in the regression equation* $\ln\{I(\lambda_{j,T})\} = \beta_0 + \beta_1 \ln\{4\sin^2(\lambda_{j,T}/2)\} + u_{j,T}$, $j = 1, \ldots, n$. *Then there exists a function* $g(T)$ (*which will have the properties* $\lim_{T\to\infty} g(T) = \infty$, $\lim_{T\to\infty} g(T)/T = 0$) *such that if* $n = g(T)$ *then* $\text{plim}\, b_1 = -d$. *If* $\lim_{T\to\infty} (\ln T)^2/g(T) = 0$, *then* $(b_1 + d)/\{\hat{\text{var}}\,(b_1)\}^{1/2} \to^{\mathcal{D}} N(0, 1)$, *where* $\hat{\text{var}}\,(b_1)$ *is the usual least squares estimator of* $\text{var}\,(b_1)$.

Once the parameter '$d$' is estimated, the relationship $f_u(\lambda) = f(\lambda)/f_2(\lambda; d)$ may be used to estimate $f_u(\lambda)$. This can be done as follows. First, the exact finite Fourier transform of the sequence $\{X_t, t = 1, \ldots, T\}$ is computed at the harmonic ordinates. The exact, rather than the fast, Fourier transform is used because the fast Fourier transform presumes circular stationarity; this presumption seems a poor approximation for long memory models, and in experiments along the lines of those reported in the next two sections the exact Fourier transform was a much more satisfactory (though more costly) procedure. Second, the Fourier transform of the series is multiplied by $(1 - \exp(-i\lambda))^{\hat{d}}$. Third, the exact inverse Fourier transform of this product is computed. In the final step conventional procedures—e.g., ARMA models, long autoregressions, or nonparametric spectral methods—are used to model $f_u(\lambda)$. Since $\hat{d}$ is consistent for $d$, the estimator (or the implicit estimator) of $f_u(\lambda)$ in the last step is consistent. The conventional distribution theory for the estimator will not be applicable, because $\hat{d}$ rather than $d$ is used in the second step; indeed, because $\text{var}\,(\hat{d}) = O(g(T)^{-1})$ and $\lim_{T\to\infty} g(T)/T = 0$, conventional standard errors in the final step could be badly misleading.

Multi-step-ahead forecasting is straightforward. With $\hat{d}$ in hand, the autoregression coefficient estimates $(a_2(j; \hat{d})$, see Section 1) for the 'long memory' portion of the series can be computed. The estimated autoregressive representation for the 'short memory' portion of the series (corresponding to $f_u(\lambda)$) can

also be computed. The convolution of the lag operators in these two representations provides the lag operator for the estimated autoregressive representation of $\{X_t\}$. Multi-step-ahead forecasts are then computed in the usual way.

It is clear that through suitable integer differencing the concept of integrated series can be extended, and these methods applied to cases $d \geq \frac{1}{2}$. The choice of 'suitable' integer involves subjective judgment and may become critical if $d$ lies exactly between two integers and theorem 2 is to be invoked. An alternative in such cases is to fractionally difference the original series as in steps one through three of the estimation procedure for $f_u(\lambda)$, using a prespecified value of $d$ in lieu of $\hat{d}$, to ensure that for the transformed series $d \in (-\frac{1}{2}, \frac{1}{2})$.

## 4. SIMULATION RESULTS

The results on estimation are asymptotic and restricted to the case $d < 0$. They leave open the questions of how large a sample is required for their reliability, how rapidly the number of ordinates used to estimate $d$, $g(T)$, should increase with sample size $T$, and whether they are valid for $d > 0$. Given the seemingly insurmountable difficulties in obtaining analytical answers to all of these questions, the most productive line of attack is to conduct some experiments with synthetic series. We report here the findings from a very limited simulation study; a more thorough investigation along these lines is warranted before these methods are used widely in empirical work.

The generation of synthetic data to mimic the realizations of long memory models has proved to be a challenge in its own right, precisely because widely separated points in the generated sample must be highly correlated. In early work, short memory approximations to long memory models were employed, with unsatisfactory results (see the discussion and citations in McLeod and Hipel, 1978, p. 497). A direct procedure (McLeod and Hipel, 1978) is to compute the pertinent $T \times T$ covariance matrix $C$ for given values of the parameter $H$ or $d$ and sample size $T$, and then compute the Cholesky decomposition $C = MM'$. Given a standardized normal sequence of synthetic random variables $\underset{\sim}{e} = (e_1, \ldots, e_T)'$ the synthetic long memory series is then $M\underset{\sim}{e}$. The computation of $M$ requires storage proportional to $T^2$ and time proportional to $T^3$, and is impractical for values of $T$ exceeding 200 or so. In a variant on this method Granger and Joyeux (1980) used the Cholesky decomposition to obtain the first 100 values of the simulated sample, and then used the truncated autoregression $\sum_{j=0}^{100} a_2(j; d)x_{t-j} = e_t$ to obtain successive values of the series.

In our simulations we employed the recursion algorithm usually attributed to Levinson (1947) and Durbin (1960), as extended by Whittle (1963); Jonas (1981) has recognized the potential of this method in the simulation of long memory time series. The Levinson–Durbin–Whittle algorithm implicitly provides the Cholesky decomposition of a Toeplitz matrix. (The distinguishing feature of a Toeplitz matrix $A$ is $a_{ij} = a_{i-j}$; the variance matrix of $T$ successive realizations of a wide sense stationary process is a Toeplitz matrix.) Given the values of the autocovariance function $R(0), R(1), \ldots, R(T)$ for a wide sense stationary process $\{X_t\}$ the algorithm provides the coefficients in the linear projection of $X_t$ on

$X_{t-1}, \ldots, X_{t-p}$ and the associated residual variance, for $p = 1, \ldots, T$. Storage requirements and computation time are proportional to $T^2$; the advantage in computation time relative to the Cholesky decomposition stems from the exploitation of the fact that the variance matrix is a Toeplitz form. (For $T = 400$, computation time with 64-bit double precision arithmetic is less than one minute using a VAX-11/780.)

Evidence on the characteristics of series generated in this way is provided in table I. For the example reported there, $T = 265$. In each of 4000 replications the first 26 values of the sample autocovariance function were computed. The average of these values over the replications is the estimated autocovariance function of the simulated data reported in table I; standard errors for these estimates are reported parenthetically. Since at least one million pairs of generated data are involved in each average, the standard errors are quite small. In no case do the estimated means differ from the population values by more than two standard errors. These results compare quite favourably with those of Granger

TABLE I

AUTOCOVARIANCES FUNCTIONS ESTIMATED FROM SYNTHETIC DATA[a]

| | $d = .25$ | | $d = .45$ | |
|---|---|---|---|---|
| Lag | True Population | Synthetic Population Estimate[b] | True Population | Synthetic Population Estimate[c] |
| 0 | 1.000 | .998 | 1.000 | 1.017 |
| 1 | .333 | .332 | .818 | .835 |
| 2 | .238 | .238 | .765 | .782 |
| 3 | .195 | .194 | .735 | .752 |
| 4 | .169 | .168 | .715 | .731 |
| 5 | .151 | .150 | .699 | .715 |
| 6 | .138 | .137 | .686 | .702 |
| 7 | .128 | .126 | .676 | .692 |
| 8 | .119 | .117 | .667 | .683 |
| 9 | .113 | .112 | .659 | .674 |
| 10 | .107 | .106 | .652 | .667 |
| 11 | .102 | .101 | .646 | .661 |
| 12 | .098 | .097 | .640 | .655 |
| 13 | .094 | .092 | .635 | .650 |
| 14 | .090 | .089 | .631 | .645 |
| 15 | .087 | .087 | .626 | .641 |
| 16 | .084 | .084 | .622 | .637 |
| 17 | .082 | .082 | .619 | .634 |
| 18 | .080 | .079 | .615 | .630 |
| 19 | .078 | .078 | .612 | .627 |
| 20 | .076 | .075 | .609 | .624 |
| 21 | .074 | .073 | .606 | .620 |
| 22 | .072 | .071 | .603 | .618 |
| 23 | .070 | .068 | .600 | .616 |
| 24 | .069 | .069 | .598 | .613 |
| 25 | .068 | .068 | .595 | .611 |

[a] See text for data generation method.
[b] Standard errors for these estimates range from .0014 to .0019.
[c] Standard errors for these estimates range from .0096 to .0097.

and Joyeux (1980, pp. 25–26), and underscore the importance of using exact methods rather than short memory approximations in the synthesis of long memory time series.

The results of the experiments conducted are provided in table II. The experiments were designed to investigate the effects of alternative $T$, of alternative true models, and of using the known variance $\pi^2/6$ of the disturbance of the regression equation (2) as an alternative to the least squares estimator $s^2$, all for $d > 0$. The rule $g(T) = T^\alpha$ ($\alpha = .5, .6, .7$) was held fixed. The fractions reported in table II can be regarded as point estimates of the difference between one and the true size of the confidence intervals in finite sample; with 300 replications, the standard error of these point estimates is about .0125.

TABLE II

FRACTION OF REPLICATIONS IN WHICH TRUE $d$ WAS OUTSIDE 95% CONFIDENCE INTERVAL FOR $d$ BASED ON THEOREM 2

300 REPLICATIONS, $g(T) = T^\alpha$

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

| Model | $d=.2, \phi=0$ | | | $d=.35, \phi=0$ | | | $d=.44, \phi=0$ | | | $d=.25, \phi=.5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | .5 | .6 | .7 | .5 | .6 | .7 | .5 | .6 | .7 | .5 | .6 | .7 |
| $T=50, \hat{\sigma}^2=s^2$ | .123 | .073 | .047 | .107 | .070 | .070 | .133 | .100 | .080 | .136 | .216 | .390 |
| $T=50, \hat{\sigma}^2=\pi^2/6$ | .063 | .057 | .030 | .033 | .027 | .040 | .067 | .057 | .067 | .070 | .113 | .356 |
| $T=100, \hat{\sigma}^2=s^2$ | .090 | .067 | .063 | .087 | .073 | .053 | .077 | .083 | .063 | .103 | .183 | .446 |
| $T=100, \hat{\sigma}^2=\pi^2/6$ | .057 | .060 | .053 | .053 | .037 | .040 | .047 | .057 | .037 | .083 | .130 | .473 |
| $T=200, \hat{\sigma}^2=s^2$ | .087 | .117 | .090 | .063 | .043 | .053 | .073 | .100 | .057 | .090 | .156 | .566 |
| $T=200, \hat{\sigma}^2=\pi^2/6$ | .067 | .080 | .100 | .023 | .037 | 053 | .050 | .057 | .043 | .050 | .117 | .526 |
| $T=300, \hat{\sigma}^2=s^2$ | .057 | .057 | .047 | .070 | .060 | .053 | .060 | .090 | .053 | .073 | .103 | .500 |
| $T=300, \hat{\sigma}^2=\pi^2/6$ | .067 | .060 | .047 | .037 | .033 | .050 | .050 | .053 | .060 | .053 | .070 | .513 |

The results for simple integrated series, in the first three panels of table II, are better than those for the general integrated series in the fourth panel. Most markedly, confidence intervals evidently are more reliable when the true variance $\pi^2/6$ is used than when the conventional regression $s^2$ is employed to construct standard errors. Second, the reliability of confidence intervals is insensitive to the size of $d$. Finally—by comparison with these patterns—the effect of sample size on the reliability of the confidence intervals is negligible; in any event a sample size of 100 or more certainly seems adequate. Overall, the experimental results strongly suggest that the asymptotic results in Theorem 2 are adequate in samples of models size when the true variance $\pi^2/6$ is used. They also support the conjecture that those results are valid for $d > 0$ as well as for $d < 0$.

For the general integrated series considered these generalizations remain true: true variance leads to a more reliable confidence interval than $s^2$, and the effect of increasing sample size is small by comparison. When $g(t) = T^{.5}$ the results are much like those for the simple integrated series, but as more ordinates are

incorporated in the log periodogram regression the reliability of the confidence intervals deteriorates. This is not surprising, since as ordinates are added the contribution of the term $\ln\{f_u(\lambda_j, T)/f_u(0)\}$ in (2) becomes non-negligible. The results in the last panel suggest that in empirical work $g(T)$ should be kept small if $d$ appears sensitive to choice of $g(T)$.

## 5. FORECASTING WITH LONG MEMORY MODELS

The development of long memory models has been motivated, in part, by the hope that through providing a better description of the low frequency portion of the spectral density than is possible with conventional models, they will also provide better several-step-ahead forecasts. It should be revealing to compare the mean square error of several-step-ahead forecasts generated by these models with those of other models, using actual time series, estimated parameter values, and out-of-sample forecasts. We undertook such a comparison, using three postwar monthly price indices.

Four different models were used to produce forecasts. The first is a simple integrated series, in which the parameter $d$ was estimated as described in section 3. The second is a general integrated series of the form

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

where $\phi(B)$ and $\theta(B)$ are polynomials of finite order. The parameter $d$ was estimated as in the first model, and the series $(1-B)^d X_t$ was formed using the frequency domain methods described in section 3. Beginning with this series standard ARMA modelling procedures were used to identify and estimate $\phi(B)$ and $\theta(B)$. We have discussed reasons why these two particular models might produce good several-step-ahead forecasts if one of the long memory models is an appropriate description of the behaviour of the series. The third model is conventional ARIMA, in which only integer differencing of the series is presumed. Since parsimoniously parameterized ARIMA models are incapable of describing the low frequency behaviour of series with long memory properties, several-step-ahead forecasts for this model should be poor if a long memory is in fact appropriate. The final model is an autoregression of order fifty, estimated by ordinary least squares. A fiftieth order autoregression can describe a long memory model better than a parsimoniously parameterized ARIMA model. The larger the value of $d \in (-\frac{1}{2}, \frac{1}{2})$ the more closely it approximates a simple integrated series, and it may approximate a general integrated series better than a mixed fractional difference and ARMA model. In practice, the profligate parameterization of this model would presumably increase mean square error due to imperfectly estimated parameters to a greater extent than is the case with the other three models.

The estimation procedure in the first two models requires that the number of ordinates, $g(T)$, used in the periodogram regressions be specified. In the absence of any previous experience with this estimator, in each case we chose the number of ordinates for which the corresponding value of $d$ minimized the mean square

error of 20-step-ahead, in-sample forecasts. In most cases the plot of mean square error against the number of ordinates was a smooth parabola with a well defined minimum, and in all cases the number of ordinates selected was between $T^{.55}$ and $T^{.6}$. Forecasts were computed by calculating the autoregressive representation implied by the estimated model, truncating at the fiftieth term, and then recursively computing $n$-step-ahead forecasts.

TABLE III

ESTIMATED AUTOCORRELATION FUNCTION FOR PRICE SERIES

| | Levels | | | Differences | | |
|---|---|---|---|---|---|---|
| Lags | Food CPI | WPI | CPI | Food CPI | WPI | CPI |
| 1 | .985 | .983 | .985 | .333 | .294 | .548 |
| 2 | .970 | .966 | .969 | .244 | .317 | .538 |
| 3 | .956 | .950 | .954 | .148 | .321 | .562 |
| 4 | .942 | .933 | .938 | .103 | .156 | .499 |
| 5 | .928 | .917 | .922 | .142 | .276 | .470 |
| 6 | .915 | .901 | .906 | .218 | .292 | .496 |
| 7 | .903 | .884 | .891 | .168 | .186 | .446 |
| 8 | .891 | .867 | .875 | .080 | .209 | .465 |
| 9 | .879 | .851 | .860 | .173 | .210 | .476 |
| 10 | .867 | .834 | .844 | .204 | .085 | .473 |
| 11 | .855 | .817 | .829 | .183 | .268 | .374 |
| 12 | .842 | .800 | .815 | .250 | .245 | .390 |
| 13 | .829 | .784 | .800 | .157 | .046 | .370 |
| 14 | .817 | .767 | .786 | .091 | .232 | .300 |
| 15 | .804 | .750 | .771 | .038 | .103 | .319 |
| 16 | .792 | .733 | .757 | .093 | .072 | .319 |
| 17 | .780 | .716 | .743 | .067 | .179 | .242 |
| 18 | .768 | .699 | .729 | .109 | .128 | .289 |
| 19 | .757 | .682 | .715 | .047 | .063 | .186 |
| 20 | .747 | .665 | .702 | .055 | .164 | .240 |
| 21 | .736 | .647 | .689 | .054 | .014 | .209 |
| 22 | .725 | .630 | .677 | .071 | .007 | .183 |
| 23 | .714 | .614 | .665 | .132 | .093 | .166 |
| 24 | .702 | .597 | .653 | .123 | .062 | .154 |
| 25 | .690 | .580 | .641 | .029 | .080 | .139 |
| 26 | .679 | .564 | .629 | .028 | .114 | .154 |
| 27 | .667 | .547 | .618 | .002 | .029 | .153 |
| 28 | .655 | .529 | .608 | .022 | .128 | .125 |
| 29 | .644 | .511 | .597 | .068 | .142 | .158 |
| 30 | .632 | .493 | .587 | .101 | .155 | .169 |
| 31 | .619 | .475 | .577 | .050 | .140 | .116 |
| 32 | .607 | .458 | .567 | .074 | .167 | .206 |
| 33 | .594 | .440 | .557 | .049 | .079 | .167 |
| 34 | .582 | .424 | .548 | .049 | .084 | .124 |
| 35 | .569 | .410 | .539 | .128 | .153 | .100 |
| 36 | .557 | .395 | .529 | .152 | .091 | .178 |
| 37 | .543 | .382 | .520 | .091 | .161 | .153 |
| 38 | .531 | .368 | .512 | .073 | .123 | .153 |
| 39 | .519 | .355 | .503 | .060 | .077 | .172 |
| 40 | .507 | .342 | .495 | .108 | .123 | .189 |

The consumer price index for food ('Food CPI', January 1947 to July 1978), the Wholesale Price Index ('WPI', January 1947 to February 1977) and the Consumer Price Index ('CPI', January 1947 to February 1976) were selected for study. This choice was motivated in part by Mandelbrot's argument (1973) that prices which are fractional Gaussian noise will remain so even with attempted arbitrage, and in part by Granger's demonstration (1980) that aggregates of first order Markov processes are (under certain assumptions about the distribution of parameters across the processes) simple integrated series. Granger and Joyeux have studied the mean square error of forecast for Food CPI from January 1947 to July 1978, using a simple integrated series model and a grid search estimator for $d$.

Sample autocorrelation functions for both levels and first differences of these series are provided in table III. For the differences, these functions exhibit the slow decay typical of long memory models but not of stable autoregressive moving average models with a few parameters. The first fifty sample autocorrelation

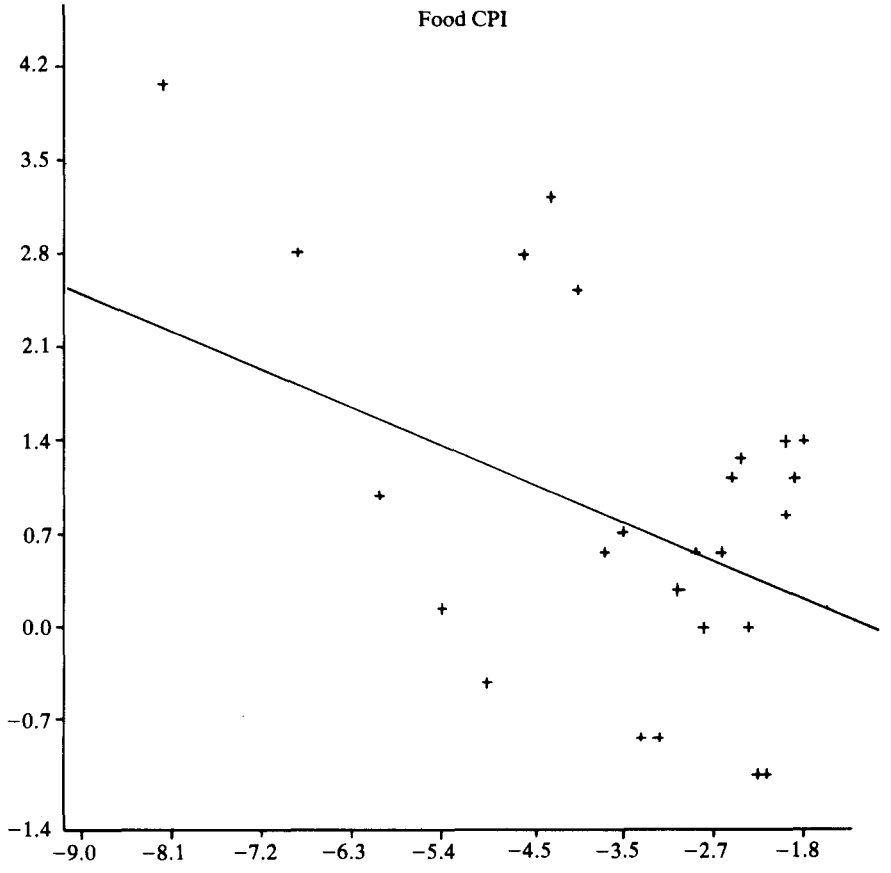


FIGURE 1. Linear plot of log periodogram vs. deterministic regressor.

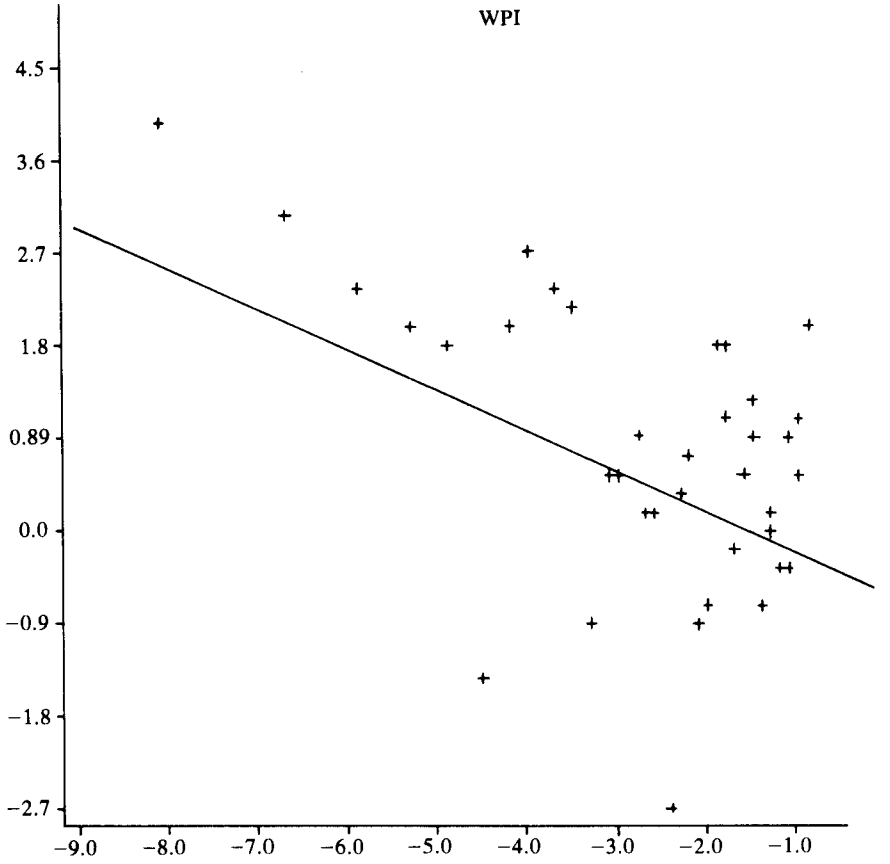


FIGURE 2. Linear plot of log periodogram vs. deterministic regressor.

coefficients are all positive in each case; for the Food CPI and the WPI over half of these are significantly different from zero by the usual rule of thumb, and for the CPI they are all significant. From table III it is clear that the CPI is more sluggish than the other two series. In the first two models, it was found necessary to begin with the series $(1-B)^{1.25}\text{CPI}_t$ (formed as described in section 3) whereas for the other two only a conventional first difference was required.

Further evidence on the long memory characteristics of these series is provided in figs. 1 through 3, in which $\ln\{I(\lambda_{j,T})\}$ (the 'dependent variable' in equation (2)) is plotted against $\ln\{4\sin^2(\lambda_{j,T})\}$ (the 'regressor' in equation (2)). For each figure the periodogram ordinates were computed from the entire sample; the number of ordinates was selected as just described. If the series in question were in fact a general fractional Gaussian noise or a general integrated series, then the relationship should be approximately linear, and the standard deviation of the plotted points about the line should be about $\pi/\sqrt{6} \doteq 1.28$. These features

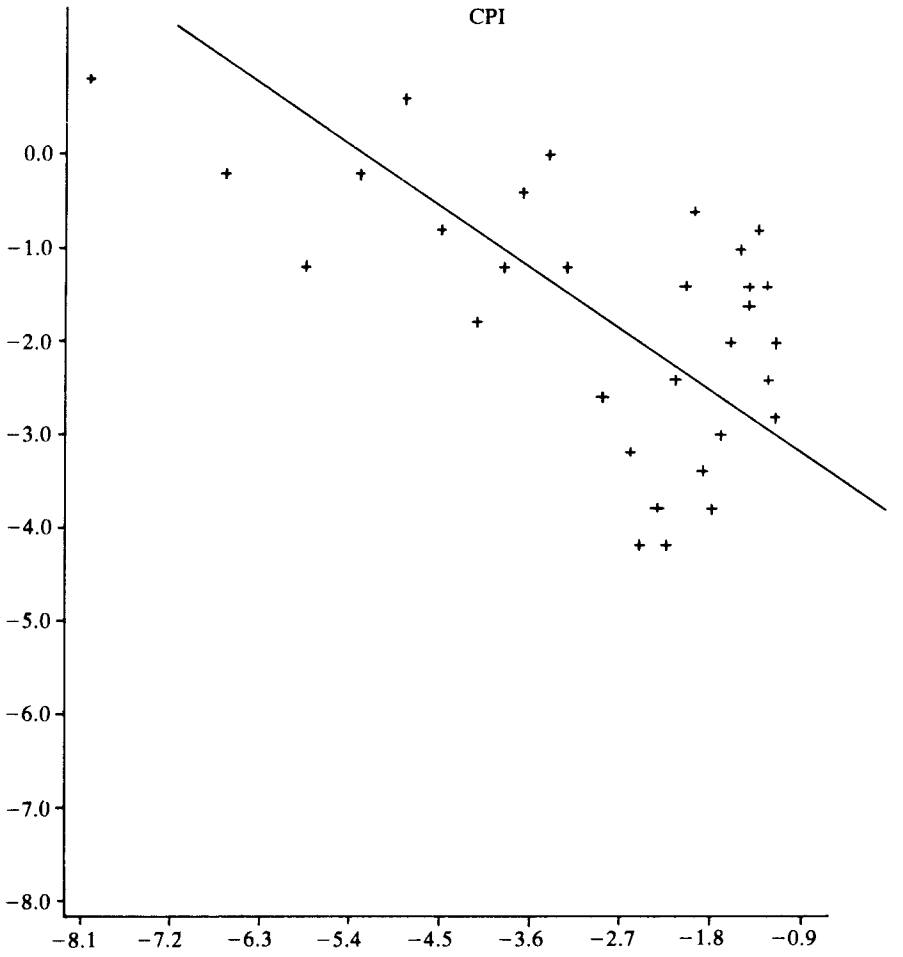


FIGURE 3. Linear plot of log periodogram vs. deterministic regressor.

seem characteristic of the plots for WPI and CPI (figs. 2 and 3), but more arguable for Food CPI (fig. 1).

For all three series 10- and 20-step-ahead forecasts were made beginning with March 1951 and at 10-month intervals thereafter. Forecasts were made using a 50 order autoregressive equation using the representation given in section 1. The parameter estimates based on the entire sample were used to produce the in-sample forecasts. For the out-of-sample forecasts parameter estimates are functions of the data preceding and including the month in which the forecast is made; the model is reestimated each time new forecasts are constructed. The mean square error of forecast is computed for the levels of the series. Results are reported in tables IV, V, and VI.

TABLE IV

COMPARISON OF MEAN SQUARE ERROR OF FORECAST CONSUMER FOOD PRICE INDEX, 1947–1978(6)

| | Mean Square Error of Forecast | | | |
|---|---|---|---|---|
| | In Sample | | Out of Sample | |
| Estimation Method | 10-Step | 20-Step | 10-Step | 20-Step |
| $d$[a] | 8.359 | 17.837 | 26.310 | 36.223 |
| $d$/ARMA[b] | 7.804 | 16.996 | 34.174 | 46.305 |
| ARMA[c] | 8.969 | 18.647 | 41.727 | 56.732 |
| AR50 | 8.046 | 16.754 | 24.017 | 33.477 |
| $\hat{d}=.35$ (Granger and Joyeux) | 8.361 | 17.923 | 28.360 | 37.443 |

[a] $\hat{d}=.423$, standard error $=.146$; $T=378$, $g(T)=26\doteq T^{.55}$.
[b] $(1-B)^d X_t=(1-.188\,B-.03\,B^2-.1\,B^3-.15\,B^4)\varepsilon_t$, where $X_t$ is the first differenced series.
[c] $(1-.33\,B)X_t=.26+\varepsilon_t$, where $X_t$ is the first differenced series.

TABLE V

COMPARISON OF MEAN SQUARE ERROR OF FORECAST WHOLESALE PRICE INDEX, 1947–1977

| | Mean Square Error of Forecast | | | |
|---|---|---|---|---|
| | In Sample | | Out of Sample | |
| Estimation Method | 10-Step | 20-Step | 10-Step | 20-Step |
| $d$[a] | 7.830 | 16.675 | 10.200 | 22.540 |
| $d$/ARMA[b] | 8.258 | 17.259 | 8.773 | 19.763 |
| ARMA[c] | 8.739 | 17.736 | 8.832 | 25.195 |
| AR50 | 9.010 | 18.030 | 38.212 | 49.242 |

[a] $\hat{d}=.417$, standard error $=.112$; $T=363$, $g(T)=39\doteq T^{.6}$.
[b] $(1-B)^d X_t=(1-.25\,B-.15\,B^2)(1+.14\,B^{12})\varepsilon_t$, where $X_t$ is the first differenced series.
[c] $(1-.18\,B-.19\,B^2-.19\,B^3-.16\,B^5)X_t=.12+\varepsilon_t$, where $X_t$ is the first differenced series.

For a given method of forecasting, it is always the case that the mean square error of the out-of-sample forecast exceeds that of the in-sample forecast, and of course 20-step-ahead forecasts are less accurate than 10-step-ahead forecasts. The comparatively sluggish CPI has a substantially lower mean square error of forecast (by whatever method) than the other two; the larger value of $\hat{d}$ for this series reflects the slower decay in the sample autocorrelation function reported in table III.

When comparisons across different forecasting procedures are made, the results are mixed but suggestive. The autoregression of order 50 ('AR50') provides the best out-of-sample forecasts for Food CPI and CPI, but performs miserably for WPI. Conventional ARMA models ('ARMA') generally perform badly: only for

TABLE VI

COMPARISON OF MEAN SQUARE ERROR OF FORECAST CONSUMER PRICE INDEX, 1947–1976

| | Mean Square Error of Forecast | | | |
|---|---|---|---|---|
| | In Sample | | Out of Sample | |
| Estimation Method | 10-Step | 20-Step | 10-Step | 20-Step |
| $d$[a] | .838 | 2.140 | 1.895 | 3.325 |
| $d$/ARMA[b] | .732 | 1.852 | 2.775 | 4.792 |
| ARMA[c] | 1.311 | 2.782 | 3.751 | 7.474 |
| AR50 | 1.402 | 2.717 | 1.417 | 3.079 |

[a] $\hat{d} = .701$, standard error = .155; $T = 351$, $g(T) = 34 \doteq T^{.6}$.
[b] $(1-B)^d X_t = (1-.5\,B)\varepsilon_t$, where $X_t$ is the first differenced series.
[c] $(1-.22\,B-.19\,B^2-.26\,B^3-.1\,B^4-.06\,B^5)X_t = \varepsilon_t$, where $X_t$ is the first differenced series.

WPI did they approach the performance of the long memory models. Mixed fractional differencing and ARMA representation ('d/ARMA') provides the best forecast for WPI, but for other series was superior only to the ARMA model. The forecasts provided by the simple integrated series model ('$d$') were never the best, but they were never markedly inferior to the others, either. The mean square error of 20-step-ahead forecasts for the simple integrated series model is on average 10% higher than that of the best set of forecasts (from whichever model). By contrast, for mixed integrated series and ARMA it is 31%, for ARMA 73%, and for an autoregression of order fifty 50%. The results suggest that use of a simple integrated series model may be a reasonable, conservative procedure for producing several-step-ahead univariate forecasts. Further experimentation along these lines with other series is clearly required to confirm this judgment.

This work was undertaken while the first author was Associate Professor and the second was a graduate student in the Department of Economics, University of Wisconsin-Madison. Portions of this article are taken from Susan Porter-Hudak's Ph.D thesis. Financial support from National Science Foundation grant SES-8207639 is acknowledged.

## REFERENCES

BUCK, R. C. (1978) *Advanced Calculus*, McGraw-Hill: New York.

DURBIN, J. (1960) The Fitting of Time Series Models. *Rev. Int. Inst. Statist.* 28, 233–244.

GRADSHTEYN, I. S. and I. M. RYZHIK (1980) *Tables of Integrals, Series and Products.* Academic Press: New York.

GRANGER, C. W. G. and R. JOYEUX (1980) An Introduction to Long Memory Time Series Models and Fractional Differencing. *J. Time Series Anal.* 1 (1), 15–29.

GRANGER, C. W. G. (1980) Long Memory Relationships and the Aggregation of Dynamic Models. *J. Econometrics* 14, 227–238.

HANNAN, E. J. (1973) Central Limit Theorems for Time Series Regression. *Zeit. Wahrscheinlichkeitsth.* 26, 157–170.

HOSKING, J. R. M. (1981) Fractional Differencing. *Biometrika* 68 (1), 165–176.

HURST, H. E. (1951) Long Term Storage Capacity of Reservoirs. *Trans. Amer. Soc. Civil Engrs.* 116, 770–799.

JONAS, A. (June 1981) Long Memory Self Similar Time Series Models. Unpublished Harvard University manuscript.

KENDALL, M. and A. STUART (1977) *The Advanced Theory of Statistics*, Vol. 1, 4th ed. Charles Griffin: London.

LEVINSON, N. (1947) The Wiener RMS Criterion in Filter Design and Prediction. *J. Math. Phys.* 25, 261–278.

MANDELBROT, B. B. and J. W. VAN NESS (1968) Fractional Brownian Motion, Fractional Noises and Applications. *SIAM Rev.* 10, 422–437.

MANDELBROT, B. B. (1971) A Fast Fractional Gaussian Noise Generator. *Water Resources Research* 7, 543–553.

MANDELBROT, B. B. (1973) Statistical Methodology for Nonperiodic Cycles: From the Covariance to R/S Analysis. *Rev. Econ. Soc. Meas.* pp. 259–290.

MCLEOD, A. I. and K. W. HIPEL (1978) Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon. *Water Resources Research* 14 (3), 491–518.

PORTER-HUDAK, S. (1982) Long-Term Memory Modelling—A Simplified Spectral Approach. Unpublished University of Wisconsin Ph.D. thesis.

ROZANOV, Y. A. (1967) *Stationary Random Processes.* Holden-Day: San Francisco.

WHITTLE, P. (1963) On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix. *Biometrika* 50, 129–134.