

Simulation and Estimation of Fractionally Integrated Time Series

JEFF BUTLER

INTERNAL REVENUE SERVICE, RESEARCH & STATISTICS OF INCOME, WASHINGTON D.C.

Fractionally integrated random processes are examined as an alternative to short-memory ARIMA models resulting from integer differencing. It is seen that fractionally integrated time series exhibit persistence at long lags of the autocorrelation function that allows for improved forecasting. Truncated and exact methods for synthetic generation are provided using SAS/IML, and a procedure for estimating fractional differencing parameter d is given using PROC SPECTRA.

1. Introduction

Following the work of Box and Jenkins (1970), it has become common practice in identifying univariate ARIMA(p, d, q) models to difference a nonstationary process Y_t exactly d times in order to achieve stationarity. When Y_t has an autoregressive representation, $\phi_p(B)Y_t = \epsilon_t$, practitioners now routinely use the results of Dickey and Fuller (1979) to test for the presence of a unit root in the polynomial $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, which implies a factorization $(1 - B)\phi'(B)$, where $\phi'(B)$ is of order $p - 1$. Such a result also implies a decomposition of Y_t into a random walk and stationary component (Beveridge and Nelson, 1981), and therefore has far-reaching consequences for long-term forecasting, as it is well-known that a random walk $Y_t = \sum_{j=0}^{\infty} \epsilon_j$ has unbounded forecast error variance as $t \rightarrow \infty$ (see for example, Nelson and Plosser, 1982).

Because of its consequences for traditional business forecasting, economists in particular have been fascinated by unit roots. As early as the 1960s, however, Mandelbrot (1968; 1969) and others began investigating the properties of fractional roots in an effort to model long-term persistence in certain river flow data, a phenomenon first observed by Hurst (1951). Early attention focused on processes represented by $(1 - B)^d Y_t = \epsilon_t$, $0 < d < 1$ and $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$. These *fractional noises* were seen to exhibit much richer dynamics in the low-frequency range of the spectrum, equivalently represented by stronger dependence at long lags of the autocorrelation function. Since then, fractionally integrated time series have been given serious attention by researchers in a wide variety of fields. However, algorithms for their use are still largely unavailable in major software packages.

This paper introduces the reader to the class of stationary univariate fractionally integrated processes and provides the essential algorithms necessary for their use in practice. Section 2 offers a theoretical foundation, while Sections 3 and 4 provide methods for simulation and estimation using the SAS® System, in particular SAS/IML and, PROC SPECTRA.

2. Statistical Background

In order to discuss the ARIMA(p, d, q) model with fractional values of d , it is necessary to understand the ARIMA(0, d , 0) process

$$(1 - B)^d Y_t = \epsilon_t$$

where $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$ is white noise. When $d = 1$, $Y_t = (1 - B)^{-d} \epsilon_t = \sum_{j=0}^{\infty} \epsilon_{t-j}$ gives the well-known random walk, but for $-1 < d < 1$, $d \neq 0$, we can use the Binomial Theorem,

$$\begin{aligned} (1 - B)^d &= \sum_{j=0}^{\infty} \binom{d}{j} (-B)^j \\ &= 1 - dB + \frac{d(d-1)}{2!} - \dots \end{aligned} \quad (1)$$

The process $Y_t = (1 - B)^{-d} \epsilon_t$, $-1 < d < 1$, $d \neq 0$ is called *fractional Gaussian noise*. Using Wold's Theorem to write $Y_t = \psi(B)\epsilon_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \dots$, it follows that when $d < 1/2$, the series Y_t is stationary with infinite moving average weights

$$\begin{aligned} \psi_k &= \frac{d(1+d) \cdots (k-1+d)}{k!} \\ &= \frac{(k+d-1)!}{k!(d-1)!} \end{aligned} \quad (2)$$

such that $\psi_k \sim k^{d-1}/(d-1)!$ as $k \rightarrow \infty$. For $d > -1/2$, Y_t is invertible with infinite autoregressive representation $\pi(B)Y_t = \epsilon_t$ and weights

$$\begin{aligned} \pi_k &= \frac{-d(1-d) \cdots (k-1-d)}{d!} \\ &= \frac{(k-d-1)!}{k!(-d-1)!} \end{aligned} \quad (3)$$

Given (2), and the fact that the autocorrelation function ρ_k for any stationary process can be written as

$\rho_k = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} / \sum_{j=0}^{\infty} \psi_j^2$, we have after some algebra

$$\rho_k = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!} \quad (4)$$

with $\rho_k \sim k^{2d-1}$ as $k \rightarrow \infty$.

For $d > 0$, induction on k shows that ψ_k and ρ_k decay hyperbolically instead of the exponential damping, $\rho_k \sim \lambda^k$, $|\lambda| < 1$ that is characteristic of ARIMA($p, 0, q$) processes. Such persistence at long lags of ρ_k is

indicative of *long memory*. Note that for $d \geq 1/2$, the power series expansion of $(1 - z)^{-d}$ is non-convergent for $|z| \leq 1$, i.e., $\lim_{k \rightarrow \infty} |\rho_k| \rightarrow 0$ but the sum $\sum |\rho_k| \rightarrow \infty$ and thus diverges; hence, $(1 - B)^d Y_t = \epsilon_t$ for $d \geq 1/2$ is a nonstationary process. Furthermore, when $d = 0$, the process is memoryless ($\rho_k = 0, |k| \geq 1$), and when $d < 0$, Y_t will exhibit short memory.

Figures 2.1 through 2.3 visually reveal the nature of the ARIMA(0, d , 0) process for $d = 0, d = 0.2$, and $d = 0.45$. Because of the influence of long lags Y_{t-k} on Y_t , the process for $0 < d < 1/2$ appears to exhibit spurious periodicity or cycles. However, there is no systematic component in these series; they all represent a 3-parameter $(\mu, \sigma_\epsilon^2, d)$ sequence of fractional noise.

Figure 2.1 ARIMA(0, 0, 0) process, or white noise.

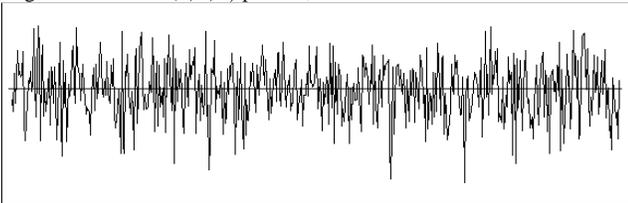


Figure 2.2 ARIMA(0, 0.2, 0) process.

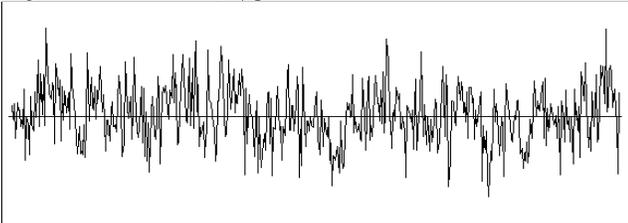
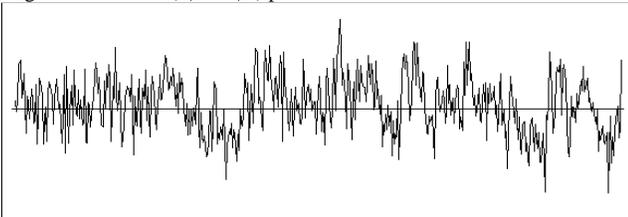


Figure 2.3 ARIMA(0, 0.45, 0) process.



One can also examine the ARIMA(0, d , 0) process by way of spectral analysis. When $d = 1$, we have a model based on first differences, $(1 - B)Y_t = \epsilon_t$, which would apply the filter $(1 - e^{-i\omega})(1 - e^{i\omega}) = |1 - e^{-i\omega}|^2$ to its spectrum, $f_Y(\omega)$. Since ϵ_t is white noise, it has spectrum $f_\epsilon(\omega) = \sigma_\epsilon^2/2\pi$, and thus for fractional d we have

$$f_Y(\omega) = (2\pi)^{-1} |1 - e^{-i\omega}|^{-2d} \sigma_\epsilon^2, \quad \omega \neq 0$$

and it is seen that $f_Y(\omega) \sim \omega^{-2d}$ as $\omega \rightarrow 0$. Recalling the relationship between $f_Y(\omega)$ and autocovariances $\gamma(k) = \rho_k \sigma_Y^2$,

$$f_Y(\omega) = \int_{-\infty}^{\infty} \gamma(k) e^{-i\omega k} dk$$

and

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k} f_Y(\omega) d\omega$$

it is obvious that a direct correspondence exists between the persistent long-lag correlations ρ_k discussed above and the rich dynamics in the spectrum of $f_Y(\omega)$ at low frequencies.

Because of the potential consequence to model specification of the choice of d , it is interesting that fractional differencing was never included in the early discussions of trend-removal methods and their implications (see Chan, Hayya, and Ord, 1977; Nelson and Kang, 1981; and Watson, 1986). For example, suppose Y_t in levels has $1/2 < d < 1$ but the researcher incorrectly takes first differences $(1 - B)Y_t = u_t$, leaving a factor $(1 - B)^{d-1}u_t = \epsilon_t$. It follows from the above that the spectrum of u_t is

$$f_u(\omega) = |1 - e^{-i\omega}|^{2(1-d)} f_\epsilon(\omega)$$

with the result being $f_u(\omega) \rightarrow 0$ as $\omega \rightarrow 0$. This situation is shown in Figures 2.4 and 2.5 below, which present simulated spectra for an ARIMA(0, 0.8, 0) process that has been correctly differenced and one that has been misspecified by assuming $(1 - B)Y_t = \epsilon_t$, i.e., $d = 1$. The implication of 'zapping out' spectral power as $\omega \rightarrow 0$ is thus of some practical importance, since this result is equivalent to misspecifying the autocorrelations ρ_k and thus moving average weights ψ_k in (2), the immediate

Figure 2.4 Spectrum $f_Y(\omega)$ after correct differencing.

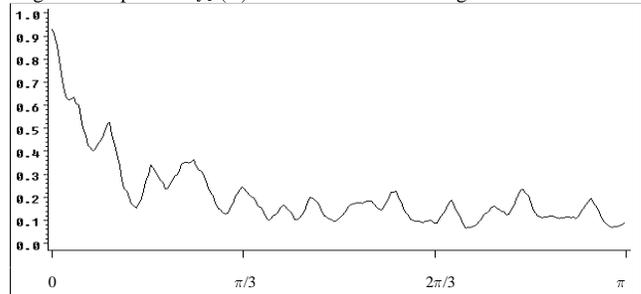
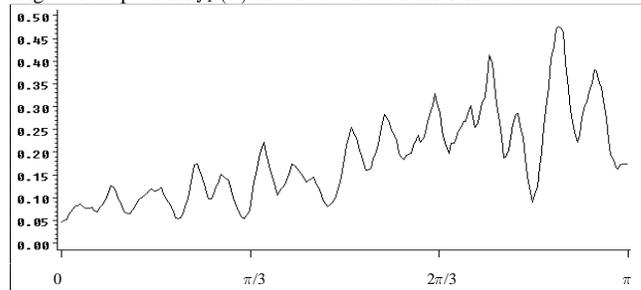


Figure 2.5 Spectrum $f_Y(\omega)$ that has been overdifferenced.



consequence of which is relatively higher root mean square forecast errors at long horizons (see Butler, 1992).

3. Simulation

There are several ways to synthetically generate an ARIMA(p, d, q) process, $-1/2 < d < 1/2$, and this paper will focus on stationary long-memory series for which $d \in (0, 1/2)$. The model is

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B)\epsilon_t \quad (5)$$

where it is assumed that y_t is stationary, i.e., the roots of the polynomials $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p) = 0$ and $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0$ lie outside the unit circle, and $0 < d < 1/2$. Since $y_t = \psi(B)\epsilon_t$, $\psi(B) = (1 - B)^{-d} \phi_p^{-1}(B)\theta_q(B)$, the problem reduces to one of efficiently solving for the convolution in B .

The fastest method for generating y_t is a two-step approximation based on a truncation of the expansion $(1 - B)^d$ either in the time or frequency domain. Let $u_t = (1 - B)^{-d}\epsilon_t$, $y_t = \phi^{-1}(B)\theta(B)u_t$. Then in the first step, one would generate the ARIMA(0, d , 0) process $u_t = (1 - B)^{-d}\epsilon_t = \psi(B)\epsilon_t$, with truncated moving average weights $\psi_k, k = 1, \dots, m$ given by (2). In the second step, one would form $y_t = \phi^{-1}(B)\theta(B)u_t$ in the usual way. Note that (2) can be rewritten as

$$\psi_k = \sum_{k=1}^m \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} B^k \quad (6)$$

where $\Gamma(x) = \int_0^\infty \lambda^{x-1} e^{-\lambda} d\lambda$ represents the gamma function, and for integers $x > 0$, $\Gamma(x) = (x-1)!$. This function is generally available in most software packages. Although it is easy to implement in practice, the one drawback of this method is that start-up values $u_{-t}, u_{-t+1}, \dots, u_0$ do not themselves belong to the fractional domain, and hence may introduce a bias in u_t . Moreover, any bias from the truncation of $\psi(B)$ will be propagated through y_t as well. This algorithm is shown in Figure 3.1, which generates a sample of $T = 250$ observations from an ARIMA(1, d , 0) process with $t = 150$ start-up values.

An exact method for simultaneously computing all parameters in $\psi(B) = (1 - B)^{-d} \phi^{-1}(B)\theta(B)$ is also available, but as with the two-step procedure above requires the truncation of $\psi(B)$ as well. Consider the power series $z = \sum_{k=1}^\infty a_k x^k$ as a function of three

parameters, α, β , and ξ . It can be shown that a recursion of a_k exists such that $(k+1)(k+\xi)a_{k+1} = (k+\alpha)(k+\beta)a_k$, which results in the *hypergeometric series*:

$$1 + \frac{\alpha}{1!} \frac{\beta}{\xi} x + \frac{\alpha(\alpha+1)}{2!} \frac{\beta(\beta+1)}{\xi(\xi+1)} x^2 + \dots$$

which will be denoted by $F(\alpha, \beta, \xi, x)$. So, for example, the stationary AR(1) process $\phi_1(B)y_t = \epsilon_t$ has a

Figure 3.1 ARIMA(1, d , 0) process with $\phi = 0.5$ and $d = 0.2$ using truncation of $\psi(B) = (1 - B)^{-d}$.

```
/* Set up arrays to store y_t, u_t, and epsilon_t. Use constants g1, g2, and g3 to define values for psi(B), and set the truncation m to 100 terms. */
```

```
data temp(keep=y251-y500);
array e{*} e1-e500;
array u{*} u1-u500;
array y{*} y1-y500;
retain d 0.2 phi 0.5 m 100;
g2=gamma(d);
do t=1 to dim(y);
  e{t}=ranncor(0);
  if t<=m then do;
    u{t}=e{t}; y{t}=u{t};
  end;
else do;
  u{t}=e{t};
  do k=1 to m;
    /* Terms in (1-B)^-d */
    c=gamma(k+d)/(g2*gamma(k+1));
    u{t}=u{t}+c*e{t-k};
  end;
  /* Final values y1, y2, ..., yT */
  y{t}=phi*y{t-1}+u{t};
end;
end;
```

convergent expansion $\psi_k = 1 + \phi_1 + \phi_1^2 + \dots$ given by $F(1, \beta, \phi, \phi) = 1/(1 - \phi)$ for any arbitrary $\beta \neq 0$.

Using the previous notation, let $u_t = (1 - B)^{-d}\epsilon_t$ be ARIMA(0, d , 0) with autocorrelation function $\rho_u(k)$, and define $y_t = \phi^{-1}(B)\theta(B)u_t = \sum_{k=0}^\infty \psi_k \epsilon_{t-k}$. Then Hosking (1981) used the hypergeometric series to show that for an ARIMA(1, d , 0) process

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} F(1, -k, 1-d-k, \phi).$$

Defining the terms $F_1 = F(1, d+k, 1-d+k, \phi)$ and $F_2 = F(1, d-k, 1-d-k, \phi)$, it can also be shown that

$$\rho_y(k) = \rho_u(k) \frac{F_1 + F_2 - 1}{(1 - \phi)F(1, 1+d, 1-d, \phi)}.$$

This function is thus not only useful for generating synthetic realizations of y_t but also for comparing the hyperbolic decay of ρ_k^d for an ARIMA(1, d , 0) process, $0 < d < 1/2$, with that of a simple AR(1) with $\rho_k = \phi^k$, as shown in Table 3.1 below.

Table 3.1 Autocorrelations of an ARIMA(1, d , 0) process with $d = 0.2$ and $\phi = 0.45$ and an AR(1) process with $\phi = 0.738$.

k	ρ_k^d	ϕ^k	k	ρ_k^d	ϕ^k
1	0.738	0.738	7	0.181	0.199
2	0.505	0.545	8	0.164	0.088
3	0.368	0.402	9	0.152	0.065
4	0.287	0.297	10	0.141	0.048
5	0.236	0.219	15	0.110	0.010
6	0.203	0.162	20	0.092	0.002

Figure 3.2 ARIMA(1, d , 0) process with $\phi = 0.5$ and $d = 0.2$ using a hypergeometric series and truncation of ψ_k .

```
/* Set truncation of (1-B)^-d to m=100 terms and F(.) to n=50 terms. */
data temp(keep=y251-y500);
array e{*} e1-e500;
array y{*} y1-y500;
```

```

retain phi 0.5 d 0.2 m 100 n 50;
g2=gamma(d);
do t=1 to dim(y);
  e{t}=rannor(0);
  if t<=k then do;
    y{t}=e{t};
  end;
  else do;
    y{t}=e{t};
    do k=1 to m;
      /* Terms in (1-B)^{-d} */
      x=1; f=1; nm=1; dm=1;
      c=gamma(k+d)/(g2*gamma(k+1));
      do j=1 to n;
        /* Terms in F(.) */
        x=x*phi;
        nm=nm*(-k+j-1);
        dm=dm*(-d-k+j);
        f=f+(nm/dm)*x;
      end;
      /* Final values y_1, y_2, ..., y_T */
      y{t}=(c*f)*e{t-k};
    end;
  end;
end;
run;
    
```

An example of using this method to generate a stationary ARIMA(1, d , 0) process is given in Figure 3.2 above. Sowell (1992) has also used this approach in deriving efficient maximum likelihood estimators.

McLeod and Hipel (1978) give an exact method for generating a nonstationary series Y_t with $1/2 < d < 1$. Mandelbrot (1969) shows that, for $0 < d < 1$, the autocorrelation function for Y_t is

$$\rho_k = \frac{1}{2}[(k+1)^{2d} - 2k^{2d} + (k-1)^{2d}], \quad k \geq 1$$

which implies $\rho_k \sim k^{2d-2}$ as $k \rightarrow \infty$. Hence, for $d > 1/2$, Y_t will have long memory; for $d = 1/2$, $\rho_k = 0$, $k \geq 1$ is white noise; and for $d < 1/2$ will have short memory. Given ρ_k for $d > 1/2$, the $P \times P$ correlation matrix can be formed, i.e.,

$$C_P(d) = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{P-1} \\ \rho_1 & 1 & \cdots & \rho_{P-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{P-1} & \rho_{P-2} & \cdots & 1 \end{bmatrix}$$

and the Cholesky decomposition of $C_P(d)$ gives

$$C_P(d) = M \cdot M'$$

where $M = m_{ij}$ is the $P \times P$ lower triangular. It is well known that exact realizations of Y_t can then be generated by

$$Y_t = \mu + \left(\sum_{i=1}^t m_{ti} \epsilon_i \right) \sqrt{\gamma_0}$$

where the variance $\gamma_0 = E(Y_t^2)$ can be predefined as well as any deterministic mean μ . This algorithm is given in Figure 3.3 below and uses PROC IML.

Since this procedure will result in nonstationary realizations of Y_t for the ARIMA(0, d , 0) process, one drawback of this approach is that any Monte Carlo studies

of stationary series will require the application of $(1-B)^{1/2}Y_t = y_t$ in the time domain or $|1 - e^{-iw}|^{1/2}$ to the Fourier transform of Y_t in the frequency domain. One could also take first differences of Y_t and apply (5), since Y_t can be factored into $(1-B)^{d-1}(1-B)Y_t = \epsilon_t$.

Figure 3.3 Exact realizations of nonstationary Y_t with $d = 0.8$.

```

/* Input data set will contain n x 1 vector
   epsilon_1, epsilon_2, ..., epsilon_n, where n = T */

proc iml;
  use _last_;
  read all var(e) into e;
  n=nrow(e);
  r=j(n,n,0);      /* Correlation matrix */
  y=j(n,1,0);     /* Final values */
  d=0.8;
  h=2*d;
  do i=1 to n;
    imj=0;
    do j=i to n;
      if j=i then r[i,j]=1;
      else do; /* Upper triangular */
        imj=imj+1;
        c1=(imj+1)**h;
        c2=imj**h;
        c3=(imj-1)**h;
        r[i,j]=0.5*(c1-2*c2+c3);
      end;
    end;
  end;
  do i=1 to n; /* Lower triangular */
    do j=1 to i;
      if i^=j then r[j,i]=r[i,j];
    end;
  end;
  m=root(r); /* Factorization */
  do i=1 to n;
    do j=1 to i; /* Final values */
      y[i,1]=y[i,1]+m[i,j]*e[j,1];
    end;
  end;
quit;
    
```

Thus, by writing $(1-B)Y_t = u_t$, one could then filter the differenced series by $u_t = (1-B)^{1-d}\epsilon_t$ to obtain a series with $0 < d < 1/2$. This will be discussed further in the next section.

Finally, Granger and Joyeux (1980) propose an approximation based on the autoregression

$$y_t + \sum_{j=1}^m \phi_j y_{t-j} = \epsilon_t$$

where the ϕ_j , $j = 1, \dots, m$ are found by solving the familiar Yule-Walker equations:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_m \end{bmatrix} = C_m^{-1}(d) \cdot \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{bmatrix}.$$

The long-memory autocorrelations ρ_k are given in (4) and the $m \times m$ matrix $C_m(d)$ is formed as $C_P(d)$ above.

Because one must truncate the autoregression to m terms, however, this method will produce a bias similar to those previously discussed.

4. Estimation

Several methods are currently available for estimating the fractional differencing parameter d . Mandelbrot and Wallis (1969) estimate $H = 1/2 + d$, where H is the rescaled range exponent used initially by hydrologists to study the long-memory properties of certain river discharge phenomena. Janacek (1982) offers a frequency domain approach, while Porter-Hudak (1982) proposes an intuitive method based on spectral regression. McLeod and Hipel (1978) and Sowell (1992) derive maximum likelihood estimators, although only the latter derives a solution for joint estimation of all parameters for the model in (5). While maximum likelihood estimation may seem the most appealing, it only offers an advantage for small samples over the two-step approach of Geweke and Porter-Hudak (1983), which will be the focus of this section.

Consider a stationary ARIMA(p, d, q) process y_t , $(1 - B)^d y_t = u_t$, $u_t = \phi^{-1}(B)\theta(B)\epsilon_t$ with spectrum $f_u(\omega)$, $0 < \omega < \pi$. Then we have

$$\begin{aligned} f_y(\omega) &= |1 - e^{-i\omega}|^{-2d} f_u(\omega) \\ &= [4\sin^2(\omega/2)]^{-2d} f_u(\omega). \end{aligned}$$

Taking logarithms and adding and subtracting a term $\ln f_u(0)$ to both sides gives

$$\begin{aligned} \ln f_y(\omega) &= \ln f_u(0) - d \ln[4\sin^2(\omega/2)] \\ &\quad + \ln f_u(\omega)/f_u(0). \end{aligned} \tag{7}$$

Adding the log-periodogram of y_t , $\ln I(\omega)$, to both sides of (7) and evaluating at harmonic ordinates $\omega_j = 2\pi j/T$, $j = 0, \dots, (T - 1)/2$ gives

$$\begin{aligned} \ln I(\omega_j) &= \ln f_u(0) - d \ln[4\sin^2(\omega_j/2)] \\ &\quad + \ln I(\omega_j)/f_y(\omega_j) + \ln f_u(\omega_j)/f_u(0) \end{aligned}$$

As $\omega \rightarrow 0$, $\ln f_u(\omega_j)/f_u(0)$ becomes negligible and can be dropped when we restrict attention to ω_j at low frequencies (note that if $u_t = \epsilon_t$, $\ln f_u(\omega)/f_u(0) = 0$ for all ω). And, since it can be shown that the term $\ln I(\omega_j)/f_y(\omega)$ is independently and identically distributed across ω_j , we can use the above result in a regression of the form

$$\ln I(\omega_j) = \alpha + \beta \ln[4\sin^2(\omega_j/2)] + \nu_j, \quad j = 1, \dots, K$$

where α and β are consistent OLS estimates of $\ln f_u(0)$ and $-d$, respectively, $\nu_j = \ln I(\omega_j)/f_y(\omega)$, and K is the upper-most harmonic ordinate to be chosen such that $0 < K < g(T)$. Since it is spectral frequencies near zero we are most interested in, it would seem that the choice of K be made carefully. Porter-Hudak (1982) provides certain regularity conditions for the function $g(T)$, and Monte Carlo experiments by

Geweke and Porter-Hudak (1983), Shea (1987), and Brockwell and Davis (1987) suggest that $g(T) = T^\lambda$, $0.5 \leq \lambda \leq 0.7$ offer the best empirical results.

Because it can be shown that there exists a function $g(T)$ such that $\text{plim}_{T \rightarrow \infty}(\hat{\beta} + d) = 0$, it is clearly important to have a large sample for this estimation procedure. A small Monte Carlo experiment was conducted to study the effects of sample size for $\lambda = 0.5$, with results shown in Table 4.1 below. In each of 500 replications, the process

Table 4.1 Distribution of $-\hat{\beta}$ for $d = 0.2$ and various sample sizes, using $g(T) = T^{0.5}$

Sample Size T	OLS Estimate of d	Standard Error
50	0.1904	0.4544
100	0.1816	0.3186
250	0.1912	0.2234
500	0.2168	0.1793

$(1 - B)^d y_t = \epsilon_t$, $d = 0.20$, was generated using a truncation of (2), with the first $T^{0.5}$ periodogram ordinates used in the regression based on (7). As discussed earlier, any bias in the estimate $-\hat{\beta}$ of d is most likely caused by the approximation used to generate y_t . However, it is clear from the standard error of $\hat{\beta}$ that efficiency will be lost in smaller samples.

PROC SPECTRA can be used to compute an estimate of the periodogram $I(\omega_j)$, and PROC REG to estimate d , as shown in Figure 4.1 below. Recall

Figure 4.1 Log-periodogram regression for estimating d .

```
/* Input series Y must be stationary */
proc spectra data=_last_ p out=out;
  var y;
run;

/* Compute ln[I(omega_j)] and ln[4sin^2(omega/2)] */
data out;
  set out;
  if _n_ <= B; /* Set B = T^lambda */
  if p_01 > 0 then y = log(p_01);
  if freq > 0 then x = log(4 * sin(freq/2)**2);
run;
proc reg; /* Estimate d */
  model y = x;
run;
```

that the input series y_t for PROC SPECTRA must be stationary. Therefore, if the original series Y_t is nonstationary, it must be differenced. Suppose the researcher believes that $(1 - B)^d Y_t = y_t$, where $y_t = \phi^{-1}(B)\theta(B)\epsilon_t$ has an ARMA representation. First differencing Y_t assumes that $d = 1$, and hence $\hat{\beta} = 0$ in the log-periodogram regression. As mentioned above, however, if $1/2 < d < 1$ in levels of Y_t , then applying the filter $(1 - B)$ will result in overdifferencing, leaving a factor $(1 - B)^{d-1}$, since $(1 - B)^{d-1}(1 - B) = (1 - B)^d$. Therefore, $\hat{\beta}$ would be an estimate of $d - 1 < 0$. In this

case, the model becomes $(1 - B)^{d-1}y_t = u_t$, $u_t = \phi^{-1}(B)\theta(B)\epsilon_t$, and one would need to apply the integrating filter $(1 - B)^{1-d}$ to the first differenced series before proceeding with the identification and estimation of ARMA parameters.

References

- Beveridge, S. and C.R. Nelson (1981). A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'. *Journal of Monetary Economics*, **7**, 151-174.
- Box, G.E.P., and G.M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Butler, J. (1992). Unit Roots and Fractional Differencing: Implications for Forecasting, in *Proceedings of the Fifth Annual Federal Forecasters Conference*, Washington, DC.
- Chan, K.H., J.C. Hayya, and J.K. Ord (1977). A Note on Trend Removal Methods: The Case of Polynomial Regression versus Variate Differencing. *Econometrica*, **45**, 737-744.
- Dickey, D.A. and W.A. Fuller (1979). Distribution of Estimates for Autoregressive Time Series with Unit Root. *Journal of the American Statistical Association*, **74**, 427-431.
- 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.
- Gradshteyn, I.S. and I.M. Ryzhik (1965). *Tables of Integrals, Series, and Products* (Fourth Ed.). Academic Press, New York.
- 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**, 15-29.
- Hosking, J.R.M. (1981). Fractional Differencing. *Biometrika*, **68**, **1**, 165-176.
- Hurst, H.W. (1951). Long-Term Storage Capacity in Reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-799.
- Janacek, G.J. (1982). Determining the Degree of Differencing for Time Series via the Log Spectrum. *Journal of Time Series Analysis*, **3**, 177-183.
- Li, W.K., and A.I. McLeod (1986). Fractional Time Series Modelling. *Biometrika*, **73**, 217-221.
- Mandelbrot, B.B. and J.W. Van Ness (1968). Fractional Brownian Motions, Fractional Noises, and Applications. *S.I.A.M. Review*, **10**, 422-437.
- Mandelbrot, B.B. and J.R. Wallis (1969). Computer Experiments with Fractional Gaussian Noises, Parts 1, 2 and 3. *Water Resources Research*, **5**, 228-267.
- McLeod, A.I. and K.W. Hipel (1978). Preservation of the Rescaled Adjusted Range, Parts 1, 2 and 3. *Water Resources Research*, **14**, 491-518.
- Nelson, C.R. and C.I. Plosser (1982). Trends and Random Walks in Macroeconomic Time Series. *Journal of Monetary Economics*, **10**, 139-162.
- Porter-Hudak, S. (1982). Long-Term Memory Modeling—A Simplified Spectral Approach. Unpublished Ph.D. Dissertation, University of Wisconsin.
- Shea, G.S. (1987). Long-Memory Models of Interest Rates: Estimation, Forecasting, and Inference for Variance Bounds on the Interest Rate Term Structure, (Manuscript), Penn State University.
- Sowell, F. (1990). The Fractional Root Distribution. *Econometrica*, **58**, 495-505.
- Sowell, F. (1992). Maximum Likelihood Estimation of Stationary Fractionally Integrated Time Series Models. *Journal of Econometrics*, **53**, 165-188.
- Watson, M. (1986). Univariate Detrending Methods with Stochastic Trends. *Journal of Monetary Economics*, **18**, 49-75.

Contact Information

The author may be contacted at:

Jeff Butler
Internal Revenue Service
Research & Statistics of Income OP:RS:R
7701/500 N. Capitol St., NW
Washington, DC 20001
E-mail: jeff.butler@m1.irs.gov