Ch. 25 Long Memory Process

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 essential stationary process, but where the autocorrelations takes **far long** 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**.

While a considerable amount of 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 fractional differenced $I(d)$, d is a fractional number, process can be regarded as a halfway house between the $I(0)$ and $I(1)$ paradigms. One attractions of long memory models is that they implies different long run predictions and effects of shocks to conventional macroeconomics approaches.

1 Definition of Long Memory

There are several possible definitions of properties of 'long memory'. Given a discrete time series process, Y_t with autocorrelations function r_j at lag j.

Definition 1:
According to Mcleod and Hipel (1978), the process possesses long memory if the quantity

$$\lim_{T\to\infty} \sum_{j=-T}^{T} |r_j| \tag{1}$$

is not finite.

Note that for a stationary and invertible $ARMA$ process having autocorrelations

$$r_k \approx A\theta^k, \quad |\theta| < 1$$

for large k, and that these values tends to zero exponentially[1] and is hence a short memory process.

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

$$S_T = \sum_{t=1}^{T} Y_t.$$

Rosenblatt (1956) defines short dependency in terms of a process that satisfies strong mixing, so that the maximal dependence between two points of a process become trivially small as the distance between these points increase.

Definition 2:
A process Y_t can be defined as having short memory if

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

exists and is nonzero, and

$$\frac{T^{-1/2}S_{[Tr]}}{\sigma} \Longrightarrow B(r), \quad for\ all\ r \in [0,1], \tag{3}$$

where $[Tr]$ is the integer part of Tr and $B(r)$ is standard Brownian motion. This allows departure from covariance stationary, but requires the existence of moments up to a certain order.

Definition 3:
A wider definition of long memory is to include any process which possesses an autocovariance function for large k, such that

$$\gamma_k \approx \Xi(k)k^{2H-2}, \tag{4}$$

[1]therefore $\lim_{T\to\infty}\sum_{j=-T}^{T}|r_j|$ is finite.

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). H is the Hurst effect or the parameters appears in the definition of fractional Brownian motion to be discussed below.

Helson and Sarason (1967) show that any process with $H > 0$ and autocovariance function given by (4) violates the strong mixing condition, and hense is long memory or long range dependent.

[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.

2 Fractional Integration Process

As introduced above, these is a need for a family of models which have all the desirable properties of:
(a) explicitly modeling long-term persistence;
(b) being flexible enough to explain both the short-term and long-term correlation structure of a series;
(c) enabling synthetic series to easily generated from the model.

The aim of this section is to introduce a family of model which does meet these requirements, by generalizing the well-known $ARIMA(p, d, q)$ model of Box and Jenkins (1976). The generalization consists of permitting the degree of differencing d to take any real value rather than being restricted to integral values; it turns out that for a suitable values of d, specifically $0 < d \leq \frac{1}{2}$, these 'fractionally differenced' process es are capable of modeling long-term persistence.

2.1 Fractional White Noise

We formally defined an $ARFIMA(0, d, 0)$, or a **fractional white noise** process to be a discrete-time stochastic process Y_t which can be represented as

$$(1 - L)^d Y_t = \varepsilon_t, \tag{5}$$

where ε_t is a mean-zero white noise and d is possibly non-integer. The following theorem give some of the basic properties of the process, assuming for convenience that $\sigma_\varepsilon^2 = 1$.

Theorem 1:
Let Y_t be an $ARFIMA(0, d, 0)$ process.
(a) When $d < \frac{1}{2}$, Y_t is a stationary process and has the infinite moving average representation

$$Y_t = \varphi(L)\varepsilon_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k}, \tag{6}$$

where

$$\varphi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!} = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}.$$

Here, $\Gamma(\cdot)$ is a *Gamma* function. As $k \to \infty$, $\varphi_k \sim k^{d-1}/(d-1)! \equiv \frac{1}{\Gamma(d)} \cdot k^{d-1}$.

(b) When $d > -\frac{1}{2}$, Y_t is invertible and has the infinite autoregressive representation

$$\phi(L)Y_t = \sum_{k=0}^{\infty} \phi_k Y_{t-k} = \varepsilon_t, \tag{7}$$

where

$$\phi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!} = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}.$$

As $k \to \infty$, $\varphi_k \sim k^{-d-1}/(-d-1)! \equiv \frac{1}{\Gamma(-d)} \cdot k^{-d-1}$.

(c) When $-\frac{1}{2} < d < \frac{1}{2}$, the autocovariance of Y_t ($\sigma_\varepsilon^2 = 1$) is

$$\gamma_k = E(Y_t Y_{t-k}) = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)} \tag{8}$$

and the autocorrelations functions is

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}. \tag{9}$$

As $k \to \infty$, $r_k \sim \frac{\Gamma(1-d)}{\Gamma(d)} \cdot k^{2d-1}$.

Proof:
For part (a).
Using the standard binomial expansion

$$(1-z)^{-d} = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)z^k}{\Gamma(d)\Gamma(k+1)}, \quad (how?) \tag{10}$$

it follows that

$$\varphi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}, \quad k \geq 1.$$

Using the standard approximation derived from Sheppard's formula, that for large j, $\Gamma(j+a)/\Gamma(j+b)$ is well approximated by j^{a-b}, it follows that

$$\varphi_k \sim k^{d-1}/\Gamma(d) = k^{d-1}/(d-1)! \simeq Ak^{d-1} \ \ (for\ a\ given\ value\ of\ d) \tag{11}$$

for k large and an appropriate constant A.

Consider now an $MA(\infty)$ model given exactly by (11), i.e.,

$$Y_t = A\sum_{k=1}^{\infty} k^{d-1}\varepsilon_{t-k} + \varepsilon_t$$

so that $\varphi_0 = 1$. This series has variance

$$Var(Y_t) = A^2\sigma_\varepsilon^2\left(1 + \sum_{k=1}^{\infty} k^{2(d-1)}\right).$$

From the theory of infinity series, it is known that

$$\sum_{k=1}^{\infty} k^{-s} \quad converges\ for\ s > 1 \tag{12}$$

but otherwise diverges. It follows that the variance of Y_t is finite provided $d < \frac{1}{2}$, but is infinite if $d \geq \frac{1}{2}$. Also, since $\sum_{k=0}^{\infty}\varphi_k^2 < \infty$ when $d < 1/2$, the fractional white noise process is mean square summable and stationary for $d < \frac{1}{2}$.[3]

The proofs of part (b) is analogous to part (a) and is omitted.

For part (c), See Hosking (1981) and Granger and Joyeux (1980) for the proof of γ_k and r_k. It is note that

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)}}{\frac{\Gamma(d)\Gamma(1-2d)}{\Gamma(1-d)\Gamma(1-d)\Gamma(d)}} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} \simeq \frac{\Gamma(1-d)}{\Gamma(d)}k^{2d-1}. \tag{13}$$

2.1.1 Relations to the Definitions of Long-Memory Process

For $0 < d < \frac{1}{2}$, the fractionally integrated process, $I(d)$, Y_t is long memory in the sense of the condition (1), its autocorrelations are all positive ($\frac{\Gamma(1-d)}{\Gamma(d)}k^{2d-1}$) such that condition (1) is violated [4] and decay at a hyperbolic rate.

For $-\frac{1}{2} < d < 0$, the sum of absolute values of the processes autorelations tends to a constant, so that it has **short memory** according to definition (1).[5]

[3]Brockwell and Davis (1987) show that Y_t is convergent in mean square through its spectral representation.

[4]Suppose that $\sum a_n$ converges. Then $\lim a_n = 0$. See Fulks (1978), p.465.

[5]From (12), $\sum_{k=0}^{\infty}\frac{\Gamma(1-d)}{\Gamma(d)}k^{2d-1}$ converges for $1-2d > 1$ or that $d < 0$.

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.

The relation of the second definition of long memory with $I(d)$ process can be illustrated with the behavior of the partial sum S_T in (2), when Y_t is a fractional white noise as in (5). Sowell (1990) shows that

$$\lim_{T\to\infty} Var(S_T)T^{-(1+2d)} = \lim_{T\to\infty} E(S_T^2)T^{-(1+2d)} = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}.$$

Hence,

$$Var(S_T) = O(T^{1+2d}),$$

which implies that the variance of the partial sum of an $I(d)$ process, with $d=0$, grows linearly,i.e., at rate of $O(T^1)$. For a process with intermediate memory with $-\frac{1}{2} < d < 0$, the variance of the partial sum grows at a **slower rate** than the linear rate, while for a long memory process with $0 < d < \frac{1}{2}$, the rate of growth is faster than a linear rate.

The relation of the third definition of long memory with $I(d)$ process can be illustrated beginning with the definition of fractional Brownian motion.

Brownian motion is a continuous time stochastic process $B(t)$ with independent Gaussian increments. Its derivatives is the continuous-time white noise process.

Fractional Brownian motion $B_H(t)$ is a generalization of these process. The fractional Brownian motion with parameter H, usually $0 < H < 1$, is the $(\frac{1}{2} - H)th$ fractional derivatives of Brownian motion. The continuous-time fractional noise is then defined as $B'_H(t)$, the derivative of fractional Brownian motion; it may also be thought of as the $(\frac{1}{2} - H)th$ fractional derivative of the continuous time white noise,[6] to which it reduces when $H = \frac{1}{2}$.

We seek a discrete time analogue of continuous time fractional white noise. One possibility is discrete time fractional Gaussian noise, which is defined to be a process whose correlation is the same as that of the process of unit increments $\triangle B_H(t) = B_H(t) - B_H(t-1)$ of fractional Brownian motion.

[6]To see this, consider that $B'(t) = \varepsilon_t$, $B'_H(t) = \varepsilon_t^*$ and $B'_H(t) = (1-L)^{(1/2-H)}B'(t)$. The fractional white noise is therefore defined as $\varepsilon_t^* = (1-L)^{(1/2-H)}\varepsilon_t$, where ε_t is a white noise process.

The discrete time analogue of Brownian motion is the random walk, X_t defined by

$$(1-L)X_t = \varepsilon_t,$$

where ε_t is *i.i.d.*. The first difference of X_t is the discrete-time white noise process ε_t. By analogy with the above definition of continuous time white noise we defined **fractionally differenced white noise** with parameter H to be the $(\frac{1}{2}-H)th$ fractional difference of discrete time white noise. The fractional difference operator $(1-L)^d$ is defined in the natural way, by a binomial series:

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty}\binom{d}{k}(-L)^k = 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - ... \\ &= \sum_{k=0}^{\infty}\frac{\Gamma(k-d)z^k}{\Gamma(-d)\Gamma(k+1)}. \qquad (14)\end{aligned}$$

We write $d = H - \frac{1}{2}$, so that the continuous time fractional white noise with parameters H has as its discrete time analogue the process $X_t = (1-L)^{-d}\varepsilon_t$, or $(1-L)^d X_t = \varepsilon_t$, where ε_t is a white noise process.

With the results above, the fractional white $I(d)$ process is also a long memory precess according to definition 3 by substitution $d = H - \frac{1}{2}$ into (9).

2.2 ARFIMA process

A natural extension of the fractional white noise model (5) is the **fractional** *ARMA* model or the $ARFIMA(p,d,q)$ model

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, \qquad (15)$$

where d denotes the fractional differencing parameter, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$ and ε_t is white noise. The properties of an *ARFIMA* process is summarized in the following theorem.

Theorem 2:
Let Y_t be an $ARFIMA(p,d,q)$ process. Then
(a) Y_t is stationary if $d < \frac{1}{2}$ and all the roots of $\phi(L) = 0$ lie outside the unit circle.

(b) Y_t is invertible if $d > -\frac{1}{2}$ and all the roots of $\theta(L) = 0$ lie outside the unit circle.

(c) If $-\frac{1}{2} < d < \frac{1}{2}$, the autocovariance of Y_t, $\gamma_k = E(Y_t Y_{t-k}) \sim B \cdot k^{2d-1}$, as $k \to \infty$, where B is a function of d.

Proof:
(a). Writing $Y_t = \varphi(L)\varepsilon_t$, we have $\varphi(z) = (1-z)^{-d}\theta(z)\phi(z)^{-1}$. Now the power series expansion of $(1-z)^{-d}$ converges for all $|z| \leq 1$ when $d < \frac{1}{2}$, that of $\theta(z)$ converges for all z and θ_i since $\theta(z)$ is polynomial, and that of $\phi(z)^{-1}$ converges for all $|z| \leq 1$ when all the roots of $\phi(z) = 0$ lie outside the unit circle. Thus when all these conditions are satisfied, the power series expansion of $\varphi(z)$ converges for all $|z| \leq 1$ and so Y_t is stationary.

(b). The proof is similar to (a) except that the conditions are required on the convergence of $\pi(z) = (1-z)^d\phi(z)\theta(z)^{-1}$.

(c). See Hosking (1981) p.171.

The reason for choosing this family of $ARFIMA(p,d,q)$ process for modeling purposes is therefore obvious from Theorem 2. The effect of the d parameter on distant observation decays hyperbolically as the lag increases, while the effects of the ϕ_i and θ_j parameters decay exponentially. Thus d may be chosen to describe the **high-lag correlation** structure of a time series while the ϕ_i and θ_j parameters are chosen to describe the **low-lag correlation** structure. Indeed the long-term behavior of an $ARFIMA(p,d,q)$ process may be expected to be similar to that of an $ARFIMA(0,d,0)$ process with the same value of d, since for very distant observations the effects of the ϕ_i and θ_j parameters will be negligible. Theorem 2 shows that this is indeed so.

Exercise:
Plot the autocorrelation function for lags 1 to 50 under the following process:
(a) $(1-0.8L)Y_t = \varepsilon_t$;

(b) $(1-0.8L)Y_t = (1-0.3L)(1-0.2L)(1-0.7L)\varepsilon_t$;
(c) $(1-L)^{0.25}Y_t = \varepsilon_t$;
(d) $(1-L)^{-0.25}Y_t = \varepsilon_t$.

2.2.1 Calculation of the Autocovariance Functions of $ARFIMA(p,d,q)$

Sowell (1992) derives the autocovariance of the stationary $ARFIMA(p,d,q)$ process. They are complicated functions of hypergeometric functions. Theses autocovariance are needed for conducting the exact maximum likelihood estimation as well as for generating data in Monte Carlo studies,[7] both of which have been pursued quite extensively in recent years.

Theorem 3 (Sowell 1992):
Let $\rho_j,\ j=1,...,p$ be the p distinct roots of a stationary $AR(p)$ process with the p coefficients $\phi_1,\phi_2,...,\phi_p$, and define

$$\zeta_j \equiv \left[\rho_j \prod_{i=1}^{p}(1-\rho_i\rho_j)\prod_{m\neq j}^{p}(\rho_j-\rho_m)\right]^{-1}.$$

Also, consider the autocovariance of an invertible $MA(q)$ process with the q coefficients $\theta_1,\theta_2,...,\theta_q$;

$$\varphi_i = \sum_{k=0}^{q-|i|}\theta_k\theta_{k+|i|},\quad i=0,1,...,q,$$

and the autocovariance of a fractional white noise process (assume $\sigma_\varepsilon^2=1$):

$$\gamma_k = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)},\quad k=0,1,2,.... \tag{16}$$

Then the autocovariance of the $ARFIMA(p,d,q)$ process as in (15) can be expressed as

$$\gamma_k^* = \gamma_k\sum_{j=1}^{p}\zeta_j A(k,\rho_j),\quad k=0,1,2,..., \tag{17}$$

where

$$A(k,\rho_j) = \sum_{i=-q}^{q}\varphi_i B(k,p+i)[\rho_j^{2p}C(p+i-k,\rho_j)+C(k-p-i,\rho_j)-1], \tag{18}$$

[7]See section 2.2.3 below.

$$B(k,h) = \frac{\Gamma(1-d+k)\Gamma(d+k+h)}{\Gamma(d+k)\Gamma(1-d+k+h)},$$

and

$$C(h,\rho_j) = \sum_{s=0}^{\infty} \frac{\Gamma(1-d+h)\Gamma(d+h+s)}{\Gamma(d+h)\Gamma(1-d+h+s)}\rho_j^s.$$

As we seen the Sowell's original formula lies in the evaluation of a large number of hypergeometric functions. Obviously, an efficient algorithm for calculating such autocovariances is called for. Chung (1994) provides some alternative methods for the calculation of these autocovariances. We present it in the following.

Theorem 4 (Chung 1994):
The expression for $A(k,\rho_j)$ in (18) can be expressed alternatively as:

$$A(k,\rho_j) = \rho_j^p \sum_{i=0}^{q} \varphi_i \alpha_i(k,\rho_j), \quad k = 0,1,2,...,$$

where

$$\alpha_i(k,\rho_j) = (\rho_j^i + \rho_j^{-i})D(k,\rho_j) + \sum_{s=1}^{p+1-i} (\rho_j^{i-s} - \rho_j^{s-i})B(k,s) + \delta_i(k,\rho_j),$$

with

$$\delta_i(k,\rho_j) = \begin{cases} -D(k,\rho_j), & if \quad i=0 \\ \sum_{s=1}^{p-1-i}(\rho_j^{-i-s} - \rho_j^{s+i})B(k,s), & if \quad 0<i<p-1 \\ 0, & if \quad 0<i=p-1 \\ \rho_j^p - \rho_j^{-p}, & if \quad i=p \\ \rho_j^i - \rho_j^{-i} + \sum_{s=1}^{i-p}(\rho_j^{i-s} - \rho_j^{s-i})B(-k,s), & if \quad i>p, \end{cases}$$

and

$$D(k,\rho_j) = C(k,\rho_j) + C(-k,\rho_j) - 1.$$

2.2.2 Mean-Reverting of $ARFIMA(p,d,q)$

In the study of unit root process, the term 'nonstationary' seems to have the same meaning with a process that have permeant effect from shock.[8] However,

[8] See section 2.2.3 of Chapter 19.

in this section, we show that **this is not the case**. An innovation may have no long-run impact even on a nonstationary fractionally integrated process.

Consider again the $ARFIMA(p, d, q)$ process

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, \tag{19}$$

where all the roots of $\phi(L) = 0$ and $\theta(L) = 0$ lie outside the unit circle and ε_t is $i.i.d.(0, \sigma_\varepsilon^2)$.

Model (19) includes $I(1)$; that is, $d = 1$, as a special case. The distinction between $d = 1$ and $d < 1$ is crucial in terms of **mean-reversion** property of Y_t. Although the effect of any shock (ε_t) is known to persist forever for an $I(1)$ process, it dies out, albeit slowly, for an $I(d)$ process with $d < 1$, where Y_t is nonstationary when $\frac{1}{2} < d < 1$ from Theorem 1. This can be seen by studying the moving average representation for $(1-L)Y_t$:

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

where $A(L) = 1 + A_1 L + A_2 L^2 + ...$, derived from

$$A(L) = (1-L)^{1-d}\psi(L), \tag{20}$$

with $\psi(L) = \phi(L)^{-1}\theta(L)$. The impact of a unit innovation at time t on the value of Y at time $t+k$ is equal to $1+A_1+A_2+...+A_k$.[9] For a mean-reverting process, the infinite impulse response $A(1)$ $(= 1 + A_1 + A_2 + ...)$ equals 0, implying no long-run impact of the innovation on the value of Y. Using (10) to find the series representation for $(1-L)^{1-d}$, (20) can be written as

$$A(L) = F(d-1, 1, 1; L)\psi(L), \tag{21}$$

where $F(\cdot)$ is the hypergeometric function defined by

$$F(m, n, p; L) = \frac{\sum_{k=0}^{\infty} \Gamma(m+j)\Gamma(n+k)\Gamma(p)L^k}{\Gamma(m)\Gamma(n)\Gamma(p+k)\Gamma(k+1)}.$$

Using some known properties of the hypergeometric function (Gradshteyn and Ryzhik 1980, pp. 1039-1042), it can be shown that $F(d-1, 1, 1, ; 1) = 0$ for $d < 1$. It follows that

$$A(1) = F(d-1, 1, 1, ; 1)\psi(1) = 0$$

[9] See also section 2.2.3 of Chapter 19.

for $d < 1$.[10] Hence an $I(d)$ process with d less than unity is **mean-reverting**.

Note that when $\frac{1}{2} \leq d < 1$, the Y_t process is covariance nonstationary because its variance is not finite.[11] Nonetheless, the Y_t process is mean-reverting, since an innovation has no permanent effect on the value of Y_t. This is in contrast to an $I(1)$ process, which is both **covariance nonstationary** and **not mean-reverting**. For an $I(1)$ process, the effect of an innovation can persistent forever.

2.2.3 How to Simulate an $ARFIMA(p, d, q)$ Process

A sample of size T for the $ARFIMA(p, d, q)$ process (15), $\mathbf{y}$, can be formed as follows.[12] First, a $(T \times 1)$ vector, $\mathbf{v}$, consisting of $N_T(\mathbf{0}, \mathbf{I}_T)$ is generated from *Gauss*. Then the desired $T \times T$ covariance matrix $\mathbf{\Sigma}$ is constructed. This is simply the Toeplitz matrix formed from the autocovariances, which is given by (17).[13] We next obtain the Choleski factorization of $\mathbf{\Sigma}$, $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$, where $\mathbf{P}$ is lower triangular. Finally the sample, $\mathbf{y}$, is generated as $\mathbf{y} = \mathbf{P}\mathbf{v}$; clearly $Cov(\mathbf{y}) = \mathbf{P}\mathbf{P}' = \mathbf{\Sigma}$. Construction of $\mathbf{y}$ in this way eliminates dependence on pre-sample startup values, which can be particularly problematic with long-memory models.

[10] Can this result be look apparent form (11) that when $d < 1$, $\lim_{j\to\infty} \varphi_j = 0$?

[11] The fact is from its infinite moving average coefficients are not square-summable.

[12] In fact, this method is applicable to the simulation of any sample for the stationary $ARMA$ model.

[13] The elements of an $m \times m$ Toeplitz matrix $\mathbf{A}$ satisfy $a_{ij} = a_{j-i}$ for scalars $a_{-m+1}, a_{-m+2}, ..., a_{m-1}$; that is, $\mathbf{A}$ has the form

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdot & \cdot & \cdot & a_{m-2} & a_{m-1} \\ a_{-1} & a_0 & a_1 & \cdot & \cdot & \cdot & a_{m-3} & a_{m-2} \\ a_{-2} & a_{-1} & a_0 & \cdot & \cdot & \cdot & a_{m-4} & a_{m-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{-m+2} & a_{-m+3} & a_{-m+4} & \cdot & \cdot & \cdot & a_0 & a_1 \\ a_{-m+1} & a_{-m+2} & a-m+3 & \cdot & \cdot & \cdot & a_{-1} & a_0 \end{bmatrix}.$$

2.3 Fractional Brownian Motion

Mandelbrot and Van Ness (1968) introduced fractional Brownian $B_d(r)$, for $d \in (-\frac{1}{2}, \frac{1}{2})$ and $0 \leq r \leq 1$:

$$B_d(r) \equiv \frac{1}{\Gamma(1+d)V_d^{\frac{1}{2}}}\left(\int_0^r (r-x)^d dB(x) + \int_{-\infty}^0 [(r-x)^d - (-x)^d]dB(x)\right). \quad (22)$$

Here, $B(r)$ is the standard Brownian motion and

$$V_d \equiv \frac{1}{\Gamma(1+d)^2}\left(\frac{1}{1+2d} + \int_0^\infty [(1+\tau)^d - \tau^d]^2 d\tau\right) = \frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)},$$

this scale constant being chosen to make $E(B_d(1))^2 = 1$. It is easily verified that $E(B_d(r)) = 0$.[14] Note that $B_d(r) = B(r)$ when $d = 0$. For $d \neq 0$, $B_d(r)$ can be formally interpreted as a fractional derivative of $B(r)$ in the sense of Weyl (1917).

A fractional Brownian motion differs from a standard Brownian motion $B(r)$ by having **correlated increments**, positively correlated when $d > 0$ and negatively correlated otherwise. Thus, it is easily verified from the definition in (22) that

$$E(B_d(r+\delta) - B_d(r))^2 = \delta^{2d+1}$$

for $r \in [0, 1)$ and $0 < \delta < 1 - r$ and hence that, for example,

$$E[B_d(r)(B_d(r+\delta) - B_d(r))] = \frac{1}{2}((r+\delta)^{2d+1} - r^{2d+1} - \delta^{2d+1}), \quad (how?) \quad (23)$$

when $d = 0$, this result is back to standard Brownian motion which has independent increment. (See p.6 of Chapter 21.)

2.4 FCLT of Fractional Integrated Process

It is often possible to characterize limiting distributions of discrete stochastic processes as functions of continuous stochastic processes by applying functional central limit theorems. The functional central limit theorems that apply to fractionally integrated times series are presented in Davydov (1970), Avram and Taqqu (1987), Mielniczuk (1997), Davidson and DeJong (2000), Chung (2002)

[14] This is what Marinucci and Robinson (1999) called 'Type I' fractional Brownian motion. There is another form of definition that is called 'Type II' fractional Brownian.

and Wang, Lin, and Gulati (2003) et.al. and the continuous stochastic process is the fractional Brownian motion. We present three of the applicability of these functional central limit theorems to fractionally integrated series in the followings.

Theorem 5 (Davydov 1970, Fractional (differenced) *i.i.d.* process):
A fractional white noise process Y_t, is written as

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

where ε_t is *i.i.d.* with zero mean and variance σ_ε^2 and $E|\varepsilon_t|^r < \infty$ for $r \geq max[4, -8d/(1+2d)]$. Define the variance of the partial sum of Y_t by $\sigma_T^2 = Var(\sum_{t=1}^T Y_t)$. Then for $-\frac{1}{2} < d < \frac{1}{2}$,

$$B_{d,T}(r) = \frac{\sum_{t=1}^{[Tr]^*} Y_t}{\sigma_T} \Rightarrow B_d(r).$$

This is a generalization of Donsker Theorem. When $d = 0$, $\sigma_T^2 = T\sigma_\varepsilon^2$, and then

$$B_T(r) = \frac{\sum_{t=1}^{[Tr]^*} Y_t}{T^{1/2}\sigma_\varepsilon} = \frac{T^{-1/2}\sum_{t=1}^{Tr} Y_t}{\sigma_\varepsilon} \Rightarrow B(r)$$

as (10) of Chapter 21.

As McLeish has extended Donsker's Theorem to the case that the innovation is a mixing process, Davidson and DeJong (2000) extend the Davydov's FCLT results to the case where the innovation after d differencing is a mixing process. We present them in the following.

Theorem 6 (Davidson and DeJong 2000, Fractional (differenced) mixing process):
Let the fractionally integrated Y_t be

$$(1-L)^d Y_t = u_t,$$

where u_t satisfies the following assumption:
(a) has zero mean;
(b) satisfies $\sup_t E|\varepsilon_t|^\gamma < \infty$ for some $\gamma > 2$;
(c) is stationary, [15] and $0 < \sigma_u^2 < \infty$, where $\sigma_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \sum_{s=1}^T E(u_t u_s)$;[16]

[15] This is a stronger condition than Phillips (1987)'s. This condition excludes any possible heterogeneous innovation.
[16] Hence, u_t is a short memory process according to this assumption.

(d) is strong mixing with mixing coefficients α_m that satisfy $\sum_{m=1}^{\infty} \alpha_m^{1-2/\gamma} < \infty$.

Define the variance of the partial sum of Y_t by $\sigma_T^2 = Var(\sum_{t=1}^{T} Y_t)$. Then for $-\frac{1}{2} < d < \frac{1}{2}$,

$$B_{d,T}(r) = \frac{\sum_{t=1}^{[Tr]^*} Y_t}{\sigma_T} \Rightarrow B_d(r).$$

On the other hand, Wang et.al. (2003) extend Davydov's FCLT results parametrically. In particular, they assume that the innovation after d differenceing is a $MA(\infty)$ process.

Theorem 7 (Wang, Lin, and Gulati 2003, Fractional (differenced) $MA(\infty)$ process):
Let the fractionally integrated Y_t be

$$(1-L)^d Y_t = v_t,$$

where v_t is a linear process of an $i.i.d.$ random variable, say η_t, i.e.

$$v_t = \sum_{j=0}^{\infty} \varphi_j \eta_{t-j}, \quad t = 1, 2, ...$$

with $\sum_{j=0}^{\infty} j^{(1/2)-d} |\varphi_j| < \infty$, $\sum_{j=0}^{\infty} \varphi_j \neq 0$ and $E|\eta_t|^{max\{2,2/(1+2d)\}} < \infty$.

Define the variance of the partial sum of Y_t by $\sigma_T^2 = Var(\sum_{t=1}^{T} Y_t)$. Then for $-\frac{1}{2} < d < \frac{1}{2}$,

$$B_{d,T}(r) = \frac{\sum_{t=1}^{[Tr]^*} Y_t}{\sigma_T} \Rightarrow B_d(r).$$

2.5 LLN and CLT of Fractional Integrated Process, See Hosking, 1996

In this section we consider an interesting question that does the laws of large number (LLN) and the central limit theorem (CLT) still holds for long memory process whose autocovariance are not absolutely summable.

Let Y_t be a covariance stationary process with mean $E(Y_t) = \mu$ and autocovariance γ_k which satisfy

$$\gamma_k = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)} \sim \lambda k^{-\alpha}, \tag{24}$$

where $\lambda = \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} > 0$, $\alpha = 1-2d$ and $0 < \alpha < 1$.

The sample mean of a realization $Y_t, t = 1, 2, ..., T$ of a times series is

$$\bar{Y} = T^{-1}\sum_{t=1}^{T} Y_t,$$

and has mean μ and variance

$$Var(\bar{Y}) = T^{-2}\sum_{t=1}^{T}\sum_{j=1}^{T}\gamma_{t-j}.$$

In the following we consider the large sample properties of the sample mean of long memory times series. We state the main limit theorems here in Hosking's (1996) form.

Theorem 8 (Hosking 1996, LLN):
Let $Y_t,\ t = 1, 2, ..., T$ be a sample of size T from a covariance stationary times series whose autocovariance function γ_k satisfies (24). Then

$$Var(\bar{Y}) \sim \frac{2\lambda T^{-\alpha}}{(1-\alpha)(2-\alpha)}.$$

Proof (Informal):
From Sowell'sa results that the variance of the partial sum of an fractional white noise is $O(T^{(1+2d)})$, then

$$\begin{aligned} Var(\bar{Y}) = Var\left(\frac{\sum Y_t}{T}\right) = \frac{T^{1+2d}\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}}{T^2} &= T^{2d-1}\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)} \\ &\equiv \frac{2\lambda T^{-\alpha}}{(1-\alpha)(2-\alpha)}. \end{aligned}$$

Although the long memory process violate the mixing condition, however from the result above, $E(\bar{Y}) = \mu$ and $Var(\bar{Y})$ is $o_p(1)$, it follows that there exist the laws of large number for the long memory (LLN) process:

$$\bar{Y} \xrightarrow{m.s.} \mu,$$

which implies

$$\bar{Y} \xrightarrow{p} \mu.$$

Theorem 9 (Hosking 1996, CLT):
In addition to (24), suppose that the time series Y_t also satisfies

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

where a_t is a white noise process consisting of independent and identically distributed random variables with second moment exists $E(a_t^2) < \infty$. We further assume that in this presentation[17]

$$\psi_j \sim \delta j^{-\beta},\ \delta > 0,\ \beta = \frac{1}{2}(1+\alpha).$$

Then the central limit theorem (CLT) for this long memory process is expressed as

$$T^{\alpha/2}(\bar{Y} - \mu) \xrightarrow{L} N\left(0, \frac{2\lambda}{(1-\alpha)(2-\alpha)}\right).$$

[17] i.e. $-\beta = -1/2 - (1/2)\alpha = -1/2 - (1/2)(1-2d) = d - 1$ as in (6).

3 Estimation and Testing

3.1 Semi-parametric Estimation of d in the Frequency

3.1.1 Spectral Analysis

Let Y_t be a covariance stationary process with mean $E(Y_t) = \mu$ and jth autocovariance γ_j. Assume that these autocovariance are absolutely summable, the autocovariance-generating function is given by

$$g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j, \tag{25}$$

where z denotes a complex scalar. If () is divided by 2π and evaluated at some z represented by $z = e^{-i\omega}$ for $i = \sqrt{-1}$ and ω a real scalar, the result is called the population spectrum of Y:

$$s_Y(\omega) = \frac{1}{2\pi} g_Y(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j}. \tag{26}$$

Note that the spectrum is a function of ω: given any particular value of ω and a sequence of autocovariance $\{\gamma_j\}_{j=-\infty}^{\infty}$, we could in principle calculate the value of $s_Y(\omega)$.

De Moivre's theorem[18] allows us to write $e^{-i\omega j}$ as

$$e^{-i\omega j} = \cos(\omega j) - i \cdot \sin(\omega j). \tag{27}$$

Substituting (27) into (26), it appears that the spectrum can equivalently be written as

$$s_Y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j [\cos(\omega j) - i \cdot \sin(\omega j)]. \tag{28}$$

Note that for a covariance-stationary process, $\gamma_j = \gamma_{-j}$. Hence, (28) implies

$$\begin{aligned} s_Y(\omega) &= \frac{1}{2\pi}\gamma_0[\cos(0) + i \cdot \sin(0)] \\ &\quad + \frac{1}{2\pi}\left\{\sum_{j=1}^{\infty} \gamma_j[\cos(\omega j) + \cos(-\omega j) - i \cdot \sin(\omega j) - i \cdot \sin(-\omega j)]\right\} \end{aligned} \tag{29}$$

[18] See Alpha Chiang (1984) p.522.

Next, we make use of the following results from trigonometry:

$$\begin{aligned}\cos(0) &= 1\\ \sin(0) &= 0\\ \sin(-\theta) &= -\sin(\theta)\\ \cos(-\theta) &= \cos(\theta).\end{aligned}$$

Using these relations, (29) simplifies to

$$s_Y(\omega) = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \cos(\omega j)\right\}. \tag{30}$$

Assuming that the sequence of autocovariances $\{\gamma_j\}_{j=-\infty}^{\infty}$ is absolutely summable, expression (30) implies that the population spectrum exists and that $s_Y(\omega)$ is a continuous, real-valued function of ω. It is possible to go a bit further and show that if the γ_j's represent autocovariance of a covariance-stationary process, then $s_Y(\omega)$ will be nonnegative for all ω. Since $\cos(\omega j) = \cos(-\omega j)$ for any ω, the spectrum is symmetric around $\omega = 0$. Finally, since $\cos[(\omega + 2\pi k)\cdot j] = \cos(\omega j)$ for any integers k and j, it follows from (30) that $s_Y(\omega + 2\pi k) = s_Y(\omega)$ for any integer k. Hence, the spectrum is a periodic function of ω. If we know the value of $s_Y(\omega)$ for all ω between 0 and π, we can infer the value of $s_Y(\omega)$ for any ω.

3.1.2 The Spectrum of a Linear Transformation

Suppose we transform Y according to

$$X_t = h(L)Y_t,$$

where

$$h(L) = \sum_{j=-\infty}^{\infty} h_j L^j$$

and

$$\sum_{j=-\infty}^{\infty} |h_j| < \infty.$$

Recall from Section 4.1 of Chapter 14 that the autocovariance-generating function of X can be calculated from the autocovariance-generating function of Y using the formula

$$g_X(z) = h(z)h(z^{-1})g_Y(z). \tag{31}$$

The population spectrum of X is thus

$$s_X(\omega) = \frac{1}{2\pi}g_X(e^{-i\omega}) = \frac{1}{2\pi}h(e^{-i\omega})h(e^{i\omega})g_Y(e^{-i\omega}). \tag{32}$$

Substituting (32) into (31) reveals that the population spectrum of X is related to the population spectrum of Y according to

$$s_X(\omega) = h(e^{-i\omega})h(e^{i\omega})s_Y(\omega). \tag{33}$$

Operating on a series Y_t with the filter $h(L)$ has the effect of multiplying the spectrum by the function $h(e^{-i\omega})h(e^{i\omega})$.

For the "difference operator" $h(L) = (1-L)$ the function $h(e^{-i\omega})h(e^{i\omega})$ would be

$$\begin{aligned} h(e^{-i\omega})h(e^{i\omega}) &= (1-e^{-i\omega})(1-e^{i\omega}) \\ &= 1-e^{i\omega}-e^{-i\omega}+1 \\ &= 2-2\cdot\cos(\omega), \end{aligned}$$

where the line follows from the fact that

$$e^{-i\omega}+e^{i\omega} = \cos(\omega)-i\cdot\sin(\omega)+\cos(\omega)+i\cdot\sin(\omega) = 2\cdot\cos(\omega).$$

To find the value of the population spectrum of X at any frequency ω, we first find the value of the population spectrum of Y vat ω and then multiply by $2-2\cdot\cos(\omega)$. For example, the spectrum at frequency $\omega = 0$ is multiplied by zero, the spectrum at frequency $\omega = \pi/2$ is multiplied by 2, and the spectrum at frequency $\omega = \pi$ is multiplied by 4. **Differencing the data removes the low-frequency components and accentuates the high-frequency components**.

3.1.3 Spectrum of an $I(d)$ process

Suppose that

$$(1-L)^d Y_t = X_t,$$

where X_t is a stationary $ARMA$ process. If X_t has spectrum $s_X(\omega)$, then Y_t does not strictly posses a spectrum,[19] but from the filtering considerations the spectrum of Y_t can be thought of as

$$s_Y(\omega) = |1-z|^{-2d} s_X(\omega), \quad \omega \neq 0,$$

where $z = e^{-i\omega}$.

3.1.4 GPH Estimator

Geweke and Porter-Hudak (1983), henceforth GPH, suggested a semiparametric estimation 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 exploit the theory of linear filters to write the process $(1-L)^d Y_t = u_t$, where $u_t \sim I(0)$, as

$$s_Y(\omega) = |1-e^{-i\omega}|^{-2d} s_u(\omega), \tag{34}$$

where $s_Y(\omega)$ and $s_u(\omega)$ are the spectral densities of Y_t and u_t respectively.

Consider a sample series of Y_t of size T. Taking logarithms of (34) and evaluating at harmonic frequencies $\omega_j = 2\pi j/T$ ($j = 0, 1, ..., T-1$), we have

$$\ln(s_Y(\omega_j)) = \ln(s_u(0)) - d\ln(4\sin^2(\omega_j/2)) + \ln[s_u(\omega_j)/s_u(0)]. \tag{35}$$

For low-frequency ordinates ω_j near 0, say $j \leq n < T$, the last term is negligible compared with the other terms. Adding $I(\omega_j)$, the periodogram at ordinate j, to both side of (36) yields

$$\ln(I(\omega_j)) = \ln(s_u(0)) - d\ln(4\sin^2(\omega_j/2)) + \ln[I(\omega_j)/s_Y(\omega_j)]. \tag{36}$$

This suggest estimating d using a simple linear regression equation

$$\ln(I(\omega_j)) = \beta_0 + \beta_1 \ln(4\sin^2(\omega_j/2))\varepsilon_j, \quad j = 1, 2, ..., n, \tag{37}$$

[19]Since a $I(d)$ process Y_t, it is long memory such that its autocovariance is not absolutely summable.

where ε_j equal $\ln[I(\omega_j)/s_Y(\omega_j)]$, is asymptotically *i.i.d.* across harmonic frequencies and $n = g(T)$ is an increasing function of T. The theoretical asymptotic variance of ε_j is known to be equal to $\pi^2/6$, which is often imposed in estimation to raise efficiency. Under some regularity conditions on $g(T)$, satisfied by, for example, T^ν for $0 < \nu < 1$, Geweke and Porter-Hudak (1983) showed that the least-square estimate of β_1 provides a consistent estimate of d and hypothesis testing concerning the value of d can be based on the t statistics of the regression coefficient.

3.2 Exact MLE Estimation of an Gaussian $ARFIMA(p, d, q)$ Model, Sowell

Consider a stationary normally distributed fractionally integrated time series Y_t generated by the model by the following $ARFIMA(p, d, q)$ model

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, \tag{38}$$

where d denotes the fractional differencing parameter, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$ and $\varepsilon_t \sim$ *i.i.d.* $N(0, \sigma^2)$. Attention will be restricted to the class of models which satisfy the following assumptions:

(a). The roots of $\phi(L) = 0$ and $\theta(L) = 0$ are outside the unit circle.
(b). $d < 1/2$,
(c). The roots of $\phi(L) = 0$ are simple.

Now let $\mathbf{y}_T$ be a sample of T observations such that $\mathbf{y}_T = [Y_1 \; Y_2 \; ... \; Y_T]'$ and $\mathbf{y}_T \sim N(\mathbf{0}, \mathbf{\Sigma})$, with logarithm of the likelihood can be expressed as

$$L(d, \phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\mathbf{y}_T'\mathbf{\Sigma}^{-1}\mathbf{y}_T,$$

where stationarity from assumption (b) implies that the covariance matrix is a Toeplitz form

$$\mathbf{\Sigma} = [\gamma_{j-i}] \;\; for \;\; i, j = 1, 2, ..., T,$$

and γ_k is the one as in (17). Sowell (1992) was able to derive the full maximum likelihood estimator for this $ARFIMA(p, d, q)$ model from this log-likelihood functions which can be evaluated on a computer. However, it is nevertheless computationally demanding, with every iteration of the likelihood requiring inversion of a T-dimensional covariance matrix and having each element a non-linear function of hypergeometric functions.

3.2.1 Asymptotic Distribution of MLE

Let $\boldsymbol{\theta} = (\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q, d)'$ denote the true parameter value where without loss of generality it is assumed that $\sigma^2 = 1$. Li and McLeod (1986) derive the asymptotic distribution of the MLE, $\hat{\boldsymbol{\theta}}$ of this $ARFIMA(p, d, q)$ model.

Theorem 10 (Li and McLeod, 1986):
The asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is normal with mean zero and covariance matrix $\mathbf{I}^{-1}$, where

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{p,q} & \mathbf{J} \\ \mathbf{J}' & \pi^2/6 \end{bmatrix}, \tag{39}$$

here, $\mathbf{J} = [\gamma_{ud}(0), ..., \gamma_{ud}(p-1), \gamma_{vd}(0), ..., \gamma_{vd}(q-1)]$, and $\mathbf{I}_{p,q}$ is the usual information matrix of the autoregressive-moving average process on $(\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q)$. Note that the information on d is independent of the $ARMA$ parameter value.

The above result holds when μ is known. When the mean μ is estimated by the maximum likelihood method the situation is more complicated and a similar result has not yet been obtained. As pointed out by the referee of Li and McLead's paper, the estimator of μ is converge at a slower rate $(T^{1/2-d}, d > 0)$ than[20] the other parameter estimator which are all $T^{1/2}$.

3.2.2 Small Sample Properties

Sufficient conditions for the consistency and asymptotic normality of the exact maximum likelihood estimators are presented in Dahlhaus (1989). However, because the fractional differencing parameter captures long-cycle characteristics of a series, asymptotic properties of MLE may be of questionable use in small samples. To discover the small sample properties, Monte Carlo simulation was used to compare different estimation procedures. Sowell finds that the MLE generally had smaller bias and MSE when the true $\mu = 0$. However, the study by Cheung and Diebold (1993) noted that the unexpected performance of Sowell's full MLE of the fractional-differencing parameter d when μ is unknown in an $ARFIMA(0, d, 0)$ model.

[20]See the results of Theorem 9.

3.3 Conditional MLE Estimation of an Gaussian $ARFIMA(p,d,q)$ Model, CSS, Chung and Baillie, 1993

Chung and Baillie (1993) consider the properties of an alternative conditional sum-of-squares (CSS) estimator of the model (38) (with a unknown μ) which minimize (see also Chapter 17, eq. (18))

$$\begin{aligned} S(d,\mu,\phi_1,\phi_2,...,\phi_p,\theta_1,\theta_2,...,\theta_q,\sigma^2) &= \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2 \\ &= \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T}[\phi(L)\theta(L)^{-1}(1-L)^d(Y_t-\mu)]^2. \end{aligned}$$

If the initial observations $Y_0, Y_{-1}, ...$ are assumed fixed, Chung and Baillie (1993) show that minimizing the conditional sum-of-square function will asymptotically equivalent to MLE. The minimum CSS procedure in the context of $ARFIMA$ processes was originally suggested by Hosking (1984). It is worth noting that similar estimation methods have been implemented in the stationary and invertible class of $ARMA$ models. For an infinite number of observations the CSS estimator will be equivalent to MLE. Similar results for $ARMA$ processes are provided by Pierce (1971).

3.3.1 Asymptotic equivalence of CSS and MLE

Chung and Baillie (1993) show setting the initial values $Y_0, Y_{-1}, ...$ to zero is immaterial in examining the asymptotic distribution of the CSS estimator in the simple fractionally integrated white-noise $ARFIMA(0,d,0)$ model. In particular they show that the asymptotic variance of CSS estimator $\hat{d}_{CSS}$ is the same as that of the MLE, which is $6/\pi^2$ as shown in (39).

3.3.2 Small Sample Properties

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 μ 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 substantially biases in a sample 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 $T = 500$. They conclude that the assumption of μ being known is far from innocuous. The estimation of μ in small sample sizes corrupts the CSS estimates of the other parameters. The resulting bias will be sufficiently large to make inference extremely unreliable.

4 Issues Relating to Unit Root and Cointegration

4.1 Power of Unit Root Test Against Fractional Alternative

See Lee and Shie (2004) for the generalized fractional unit root distribution.

4.1.1 Dickey-Fuller Test

Diebold and Rudebusch (1991) examine the properties of Dickey-Fuller unit root tests under fractionally-integrated alternatives and find that these tests have quite low power. Print p.157-158.

4.1.2 Augmented Dickey-Fuller Test

Hassler and Wolters (1994) concerned with testing the hypothesis $H_0 : \beta = 1$ in the model

$$X_t = \beta X_{t-1} + \varepsilon_t,\ (t = 1, 2, ..., T) \tag{40}$$

where the disturbance are stationary, but fractionally integrated, $(1 - L)^d \varepsilon_t = u_t, u_t \sim i.i.d.(0, \sigma^2)$. One is interested in the power of the Augmented Dickey Fuller test, i.e. the conventional t-test of $H_0; \beta = 1$ in

$$X_t = \beta X_{t-1} + \varphi_1 \triangle X_{t-1} + \cdots + \varphi \triangle X_{t-p} + \varepsilon_{tp}, \tag{41}$$

where $p \to \infty$ as $T \to \infty$.

Hassler and Wolters (1994) show that the power of this test decrease quite drastically as p increase. They also conjecture that this test is not consistent against fractional alternatives, the rationale being that, as $p \to \infty$, the ε_{tp} are approaching the independent u_t's from $(1 - L)^d \varepsilon_t = u_t$:from

$$(1 - L)^{d+1} X_t = \sum_{j=0}^{\infty} d_j X_{t-j}$$

where

$$\sum_{j=0}^{\infty} d_j = 0,\ \ d_0 = 1.$$

One can deduce the following relationships for the coefficients and disturbance in (41):

$$\beta = -\sum_{j=1}^{p+1} d_j, \quad \varphi_j = \sum_{j=i+1}^{p+1} d_j \quad and \quad \varepsilon_{tp} = u_t + \sum_{j=p+1}^{\infty} X_{t-j}.$$

Therefore, as $p \to \infty$, we have $\beta \to 1$ and $\varepsilon_{tp} \xrightarrow{p} u_t$, and one might expect that the $t-test$ of $H_0 : \beta = 1$ in (41) behaves approximately as s standard t-test does (i.e. it does not reject with increasing probability). Print p.4.

Kramer (1998) argue this intuition is misleading by showing that if p does not tend to infinity too fast, the conventional t-statistic is consistent against fractional alternatives.

4.1.3 Phillips-Perron Test

See Lee and Shie (2004).

4.2 Fractional Cointegration

In the usual framework of cointegration we have two series Y_t and X_t which are $I(1)$ and the linear combination of Y_t and X_t which is $I(0)$. However, a broader definition of fractional cointegration is that there exists an $I(d-b)$ linear combination of $I(d)$ series with $b \geq 0$. Under this definition a continuous measure of cointegration can provide more information than the $I(1)/I(0)$ framework.

In the past two decades, economists have developed a number of tools to examine whether economic variables stochastically trend together in ways that are predicted by economic theory—most notably by a cointegration test. The single equation testing procedure of Engle and Granger (1987) and the multivariate testing procedures of Johansen (1988, 1991) have become the two most popular methods of testing cointegration. They both rely on the dichotomy of a $I(1)/I(0)$ in the equilibrium error (variables that are linearly combined by a cointegrating vector), where $I(1)$ and $I(0)$ stand for the integration of order one and zero, respectively. After the appearance of a unit root[21] in all the data series is considered, Engle and Granger[22] (1987) suggest unit root tests on the residuals from the ordinary least squares estimator (OLS) of a single cointegrating vector. Phillips and Durlauf (1986) and Stock (1987) show this OLS estimator to be super consistent, while in Johansen's (1988, 1991) methodology, a Gaussian vector autoregression (VAR) is considered[23] from which one can test for the existence of one or more cointegrating vectors.

Empirical evidence such as from Diebold, Husted, and Rush (1991) and Baillie and Bollerslev (1994a) however indicate that the equilibrium error may respond more slowly to shocks, so that deviations from equilibrium are more persistent than from a stationary I(0) process. Indeed, according to Granger (1986), there is no requirement for the equilibrium error in a cointegration relationship to mimic an $I(0)$ process, as also shown by the method proposed by Engle and Granger (1987) and Johansen (1988, 1991).

Cheung and Lai (1993a) first model the equilibrium error to be a fractionally-integrated, I(d) process as introduced by Granger and Joyeux (1980) and Hosk-

[21]Banerjee et al. (1993), Stock (1994) and Phillips and Xiao (1998) survey many of the most popular methods on an unit root test.

[22]There does exist difficulties with non-standard distributions for hypothesis tests about this single cointegration vector that are due to non-zero correlation between the regressor and the disturbance. Related approaches to correct this correlation are suggested by Stock and Watson (1993) and Phillips and Hansen (1990).

[23]Lutkepohl and Claessen (1997)'s cointegrated VARMA and Saikkonen and Luukkonen (1997)'s infinite order non-gaussian VAR models are two efforts that ease Johansen's assumption.

ing (1981) in their application of testing purchasing power parity (PPP) theory. Subsequent applications of so-called fractional cointegration, (e.g., Baillie and Bollerslev (1994b), Masih and Masih (1995), Booth and Tse (1995), Hsueh and Pan (1998), Masih and Masih (1998), and Choudhry (1999) etc.) all entail a two-step testing procedure. Given that variables do share common integrated processes (usually $I(1)$), they conduct an OLS estimator of a fractional cointegrating vector and then examine whether the residuals from the OLS are $I(d)$,[24] and d is a real number less than one. The main differences in those applications are the estimation procedure of fractional difference parameter d, either by the semi-nonparametric procedure of Geweke and Porter-Hudak (1983) or the maximum likelihood estimator (MLE) of Sowell (1992), or the conditional sum of squares estimator (CSS) of Chung and Baillie (1993). However, a main drawback of those empirical applications is in the impossibility of hypothesis testing about the fractional cointegration vector that has made testing economic relationship impossible. This paper originates from this hypothesis testing problem.

To investigate the asymptotic properties of the OLS estimator of a fractional cointegration vector, Cheung and Lai (1993) show the consistency of this estimator. Tsay (2000) further derives the convergence rate of this estimator. Robinson and Marinucci (2001) first characterize the limiting distribution of both the OLS estimator and the narrow-band least square (NBLS) estimator of the fractional cointegration vector. These limiting distributions are characterized as a function of "Type II fractional Brownian Motion" (see Marinucci and Robinson (1999)). They show that the NBLS estimator demonstrates advantages over the OLS estimator in term of bias elimination and convergence rates in some cases [25]. However, under the case of fractional cointegration in a set of $I(1)$ variables with the equilibrium error being $I(d)$, $d < 1$, it is found that these two estimators make no asymptotic difference. The OLS and NBLS estimators of the fractional cointegration vector have the same convergence rate and limiting distribution.

[24] In another way, Dueker and Startz (1998) illustrate a cointegration testing methodology based on the joint estimates of the fractional orders of integration of a cointegration vector and its parent series.

[25] The cases depend on the order of integration in the regressand, regressor, and the error term. Detailed discussion of these cases could be found in Robinson and Marinucci (2001)

4.2.1 Asymptotics of OLS Estimator of Fractional Cointegration

Consider two time series , X_t and Y_t, which are $I(d)$ and are fractionally cointegrated of order (d, b)such that there exists a β that

$$Y_t = \beta X_t + \varepsilon_t,$$

where ε_t is $I(d-b)$ with $d > \frac{1}{2}$ and $d \geq b > 0$. The least squares estimator of β is given by

$$\hat{\beta} = \beta + \frac{\sum_{t=1}^{T} X_t \varepsilon_t}{\sum_{t=1}^{T} X_t^2}. \tag{42}$$

The convergence rate of $\hat{\beta}$ thus depends on theses of $\sum_{t=1}^{T} X_t \varepsilon_t$ and $\sum_{t=1}^{T} X_t^2$. This is examined in two possible situations:
(a). $(d - b \geq \frac{1}{2})$. By the Cauchy-Schwarz inequality, we have

$$\sum_{t=1}^{T} X_t^2 \sum_{t=1}^{T} \varepsilon_t^2 \geq \left(\sum_{t=1}^{T} X_t \varepsilon_t \right)^2. \tag{43}$$

This implies that

$$\left(\sum_{t=1}^{T} X_t^2 / T^{2d} \right) \left(\sum_{t=1}^{T} \varepsilon_t^2 / T^{2(d-b)} \right) \geq \left(\sum_{t=1}^{T} X_t \varepsilon_t / T^{2d-b} \right)^2. \tag{44}$$

Since it is known that[26]

$$\sum_{t=1}^{T} X_t^2 = O(T^{2d}) \ \ and \ \ \sum_{t=1}^{T} \varepsilon_t = O(T^{2(d-b)}), \tag{45}$$

Equation (44) implies that $\sum_{t=1}^{T} X_t \varepsilon_t = O(T^{\tau})$ with $\tau \leq 2d - b$. In other words, $\sum_{t=1}^{T} X_t^2 / T^{2d}$ is bounded, and $\sum_{t=1}^{T} X_t \varepsilon_t / T^{2d-b+\delta}$ converges in probability to 0 for all $\delta > 0$. It then follows from (42) that

$$T^{b-\delta}(\hat{\beta} - \beta) = \frac{\sum_{t=1}^{T} X_t \varepsilon_t / T^{2d-b+\delta}}{\sum_{t=1}^{T} X_t^2 / T^{2d}} \tag{46}$$

converges in probability to 0 for all $\delta > 0$.

(b). $(0 \leq d - b < \frac{1}{2})$. In this case

$$\sum_{t=1}^{T} X_t^2 = O(T^{2d}) \ \ and \ \ \sum_{t=1}^{T} \varepsilon_t = O(T), \tag{47}$$

[26] See Lemma A.1, (b) of Lee and shie (2004), p.297.

since ε_t is $I(d-b)$, which is stationary process for $d-b<\frac{1}{2}$. Applying the Cauchy-Schwarz inequality, (43) yield

$$\left(\sum_{t=1}^{T} X_t^2/T^{2d}\right)\left(\sum_{t=1}^{T}\varepsilon_t^2/T\right) \geq \left(\sum_{t=1}^{T} X_t\varepsilon_t/T^{d+1/2}\right)^2. \tag{48}$$

This suggests that $\sum_{t=1}^{T} X_t\varepsilon_t = O(T^\tau)$ with $tau \leq d+\frac{1}{2}$. A tighter bound, however, can be obtained by observing that

$$\sum_{t=1}^{T} X_t\varepsilon_t/T^{2d-b} \tag{49}$$

converges in distribution to some functions of Brownian motions, following from the functional central limit theorem. This implies that $\sum_{t=1}^{T} X_t\varepsilon_t = O(T^\tau)$ with $tau \leq 2d-b$, so the result in (46) still hold.

4.2.2 Empirical Application, $I(1-b)$ cases

(a). Cheung and Lai (1993):

This study examines the relevance of long-run PPP using a fractional cointegration approach that integrated the notions of cointegration and of fractional differencing. In particular, they estimated the model:

$$sp_t = \hat{\alpha}_0 + \hat{\alpha}_1 p_t + e_t,$$

where sp_t is the foreign price index converted to domestic currency units, p_t is the domestic price index, $\hat{\alpha}_i, i=1,2$ is the OLS estimators and e_t is the OLS residuals. Given that the unit root hypothesis can not be rejected on sp_t and p_t, we estimate the fractional differencing parameter of e_t. Empirical results show that all of the estimates of d lie between 0 and 1, suggesting possible fractional integration behavior of e_t, i.e., fractional cointegration behavior exists between sp_t and p_t.

(b). Baillie and Bollerslev (1994):

This study examine whether a group of exchange rate are cointegrated. In particular, they estimated the model:

$$WG_t = \hat{\alpha}_0 + \hat{\alpha}_1 UK_t + \hat{\alpha}_2 JP_t + \hat{\alpha}_3 CN_t + \hat{\alpha}_4 FR_t + \hat{\alpha}_5 IT_t + \hat{\alpha}_1 SW_t + e_t,$$

and performed the CSS to estimate the fractional differencing parameter on the cointegrating residual e_t They get an estimate of $d=0.89$ which is significantly less than one.

4.3 Regression with I(d) Regressors and Disturbance

Chung (2002) find that, even though all the regressors (one of which is $I(d_1),\ 0 < d-1 < 1/2$) and the disturbance ($I(d_2),\ 0 < d_2 < 1/2$) are stationary and ergodic, the join long memory in one single regressor and in the disturbance ($d_1+d_2 > 1/2$) can invalidate the usual asymptotic theory for OLS estimation. Specifically, the convergence rates of the OLS estimators become slower, the limits are not normal, and the standard t and F tests are all collapse.

4.3.1 MLE of Fractional Cointegration

Dueker and Startz (1998)

4.4 Fractional DF test Econometrica 2003