CHAPTER – II

CONCEPTS AND REVIEW

2.1 Introduction

The study aims at suggesting a forecasting model for the rainfed crops of Coimbatore District, Tamilnadu. The fake of the economy realize more on the stability of prices in each sector. The price also determines the cropping pattern. Hence a perfect forecaster may be a good guide for the policy makers. It is on this importance that this study was initiated. To analyse properly a knowledge on the past studies on similar topics is very essential. It is on this basis a detailed review and the definitions of concepts are presented in this chapter under two sections.

2.2 Concepts

In this section various concepts used in this study are defined on the basis of definitions found in the text books and earlier studies:

*Time series data* refers to observations on a variable that occur in a time sequence.

2.2.1 Basic Definitions

*Time series analysis* refers to any kind of analysis involving time-series data. At other times it is used more narrowly to describe attempts to explain the behavior of time-series data using only past observations on the variable in question.
A stationary series is a time-series in which the dependent variable has a constant mean and variance over time.

Parsimony is the model, which uses the smallest, number of coefficients needed to explain the available data.

Differencing is a relatively simple operation that involves calculating changes in the value of a data series.

Auto correlation function reveals how the correlation between any two values of the series as their separation changes.

Partial autocorrelation function shows the correlation between ordered pairs separated by various time spans without the effect of intervening observations accounted for. The ordered pairs are drawn from a single time series.

The basic set of relationships and the underlying process over time is referred to as the pattern in the data.

A time series $y_{t1}, y_{t2},..., y_{tn}$ describing a stochastic process is said to be stationary (or stationary in the strict sense) if the distribution of $y_{t1}, y_{t2},..., y_{tn}$ is the same as the distribution of $y_{t1+t}, y_{t2+t},..., y_{tn+t}$ for every finite set of integers $(t_1,t_2,...,t_n)$ and for every integer ‘t’.

Here $E(y_t) = m(t)$ called the mean function

$E[(y_t-m(t)) (y_s-m(s))] = Cov(y_t, y_s) = \sigma(t,s)$ called the covariance function
\( \sigma(t,s) = \sigma(s,t) \) for all \( s \) and \( t \)

A stochastic process is said to be \textit{stationary in the wider sense} or weakly stationary or stationary to second order if the mean function (or) and the variance function \( \sigma(s,t) \) satisfies the relations

\[
m(s) = m(s+t) = m
\]

and

\[
\sigma(t_1,t_2) = \sigma(t_1+t, t_2+t)
\]

setting \( t = -t_2 \), this given

\[
\sigma(t_1, t_2) = \sigma(t_1-t_2,0) = \sigma(t_1-t_2)
\]

\section*{2.2.2 Spectral distribution function and the spectral density}

A stochastic process with discrete time parameter which is stationary in the wide sense defines a sequence of covariance \( \sigma(0), \sigma(1), \ldots \). Consider the Fourier transform of the sequence,

\[
f(\lambda) = \frac{1}{2\pi} \sigma(0) + \frac{1}{\pi} \sum_{h=1}^{\infty} \sigma(h) \cos \lambda h \tag{2.2.1}
\]

\[
= \frac{1}{\pi} \sum_{h=-\infty}^{\infty} \sigma(h) \cos \lambda h, \ -\pi \leq \lambda \leq \pi
\]

multiplying (2.2.1) by \( \cos \lambda k \) and integrating from \(-\pi\) to \(\pi\)

We have
\[ \sigma(k) = \int_{-\pi}^{\pi} \cos \lambda k f(\lambda) d\lambda \]  

(2.2.2)

The function \( f(\lambda) \) is called the spectral density. \( f(\lambda) \) is even i.e. \( f(-\lambda) = f(\lambda) \),

It is convenient to treat \( g(\lambda) = 2 f(\lambda) \),

\[ 0 \leq \lambda \leq \pi \; ; \text{that is} \]

\[ g(\lambda) = \sum_{h=-\infty}^{\infty} \sigma(h) \cos \lambda h. \]

We can also call \( g(\lambda) \) as spectral density

\[ F(\lambda) = \int_{-\pi}^{\pi} f(v) dv \] is called the spectral distribution.

spectral density is often shortend to spectrum The physical meaning of the spectrum is that \( f(\lambda) d\lambda \) represents the contribution to variance of components with frequencies in the range \((\lambda, \lambda + d\lambda)\). When the spectrum is drawn it indicates that the total area underneath the curve is equal to the variance of the process. A peak in the spectrum indicates an important contribution to variance at frequencies in the appropriate region \( f(\lambda) \). It is also to be noted that the auto covariance function and spectral density function are equivalent ways of describing a stationary stochastic process. From a practical point of view, they are complementary to each other and both contain the same information expressed in different ways.

In spectral or spectrum analysis \( X_t \) is expressed as a function of cosine and sine as

\[ X_t = \mu + \alpha \cos wt + \beta \sin wt + z_t, \quad \ldots \]  

(2.2.3)
Where $z_i$ is random error term, $\mu$, $\alpha$, $\beta$ are the parameters to be estimated from the data $x_1,x_2,\ldots,x_n$ are the observed values, then the Expectation of (2.2.3) in matrix form is

$$E(X) = AQ,$$

where

$$X^T = (x_1,x_2,\ldots,x_n)$$

and

$$Q^T = (\mu, \alpha, \beta)$$

The OLS estimate of $Q$ denoted by $\hat{\theta}$

$$\hat{\theta} = (A^T A)^{-1} A^T X,$$

Now the highest frequency for this is at $\omega = \pi$ and this frequency is called the Nyquist frequency.

By equating the cycle length $\frac{2\pi}{\omega}$ to $N$ we get the lowest frequency as $\frac{2\pi}{N}$.

If $N$ is assumed to be even then

$$\omega_p = \frac{2\pi p}{N} (p=1, 2,\ldots, \frac{N}{2})$$

Now $A^T A$ becomes diagonal and $\hat{\theta}$ can be estimated.
The finite Fourier expansion of \( \{X_t\} \), is

\[
X_t = a_0 + \sum_{p=1}^{N/2-1} \left[ a_p \cos \left( \frac{2\pi pt}{N} \right) + b_p \sin \left( \frac{2\pi pt}{N} \right) \right] + a_{N/2} \cos \pi t, \quad t=1,2,\ldots,N
\]

With \( a_0 = \bar{x} = a_{N/2} = \frac{\sum (-1)^t x_t}{n} \).

\[
a_p = \frac{2\Sigma x_t \cos \left( \frac{2\pi pt}{N} \right)}{N}, \quad p=1,\ldots,\frac{N}{2} - 1
\]

\[
b_p = \frac{2\Sigma x_t \sin \left( \frac{2\pi pt}{N} \right)}{N}, \quad p=1,\ldots,\frac{N}{2} - 1
\]

Now denoting \( \frac{2\pi p}{N} \) as \( \omega_p \), the \( p^{th} \) term in the expansion can be written as: (it is also called the \( p^{th} \) harmonic)

\[
a_p \cos \omega_p t + b_p \sin \omega_p t = R_p \cos (\omega_p t + \varphi_p) \text{ (say)}
\]

where \( R_p = \sqrt{a_p^2 + b_p^2} \) called the amplitude of the \( p^{th} \) harmonic and \( \varphi_p = \tan^{-1} \left( -\frac{b_p}{a_p} \right) \) is the phase of the \( p^{th} \) harmonic. The contribution of the \( p^{th} \) harmonic to the total variability in the series is

\[
\frac{N(a_p^2 + b_p^2)}{2} = \frac{N R_p^2}{2}
\]

Now total variation can be written as

\[
\Sigma(x_t - \bar{x})^2 = N \sum_{p=1}^{n-1} \frac{R_p^2}{2} + a_{N/2}^2
\]

If we plot \( \frac{R_p^2}{2} \) against \( \omega_p = \frac{2\pi p}{N} \) we obtain a line spectrum. When \( p = N/2 \) we denote by

\[
\text{I(}\pi\text{)} = N \frac{a_{N/2}^2}{\pi}
\]
The plot of $I(\omega)$ against $\omega$ is known as periodogram. The total area under the periodogram is equal to the variance of the time series.

### 2.3 Review of Literature

We give a review of the growing literature on long memory, fractionally integrated processes that are associated, with hyperbolic decaying autocorrelations and impulse response weights. The review of the statistical literature in this field is deliberately selective.

While long memory models have only really been used by scientist since around 1980, they have played a role in the physical sciences since at least 1950, with statisticians in fields as diverse as hydrology and climatology long recognizing the presence of long memory within data recorded over both time and space. The presence of long memory can be defined from an empirical data-oriented approach in terms of the persistence of observed autocorrelations. The extent of the persistence is consistent with an essentially stationary process, but where the autocorrelations take for longer to decay than the exponential rate associated with the ARMA class. This phenomenon has been noted in different data sets by Hurst (1951, 1957), Mandelbrot and Hipel (1978) Mandelbrot (1972), and Mcleod and Hipel (1978) among others. When viewed as the time series realization of a stochastic process, the auto correlation function exhibits persistence that is neither consistent with an I (1) process nor an I(0) process. One of the most compelling motivations concerning the importance of the long memory, fractionally integrated process is related to the rate of decay associated with the impulse response coefficients of a process. The classical theory of stationary timer series, and indeed many of the models
used in econometrics, requires the existence of the wold decomposition. The conditions of the wold decomposition are relatively meak and, a part from the possible presence of a purely deterministic component, little more than square summability, and martingale behavior is required for the innovation sequence associated with the stochastic component.

A considerable amount of success has been obtained from using the ARMA class of models which impose an exponential, or geometric, rate of decay on the wold decomposition coefficients. This strategy has taken the time series analysis a long way theoretically and also in modeling empirical behavior. However, there is no conceptual reason for restricting attention to exponential rates of decay in the wold decomposition and there are indeed both theoretical and practical reasons for considering slower rates, such as hyperbolic decay. While a considerable amount of recent work has emphasized the role of persistence of shocks, most of it has been directed towards testing for the presence of unit roots in autoregressive representation of univariate and vector processes. However, the knife-edge distinction between I(0) and I(1) processes can be far too restrictive. The fractionally differentiated process can be regarded as a half-way house between the I (0) and I(1) paradigms. One attraction of long memory models is that they imply different long run predictions and effects of shocks to conventional approaches.

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

2.3.1 Preliminary ideas and definitions:

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

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

is non finite. Equivalently, the spectral density $f(w)$ will be unbounded at low
frequencies. A stationary and invertible ARMA process has autocorrelations which are
geometrically bounded, i.e. $|\rho_k| \leq C m^k$, for large $k$, $0 < m < 1$, and is hence a short
memory process. In particular, the process $y_t$ is said to be integrated of order $d$, or I($d$), if

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

Where $L$ is the lag operator, $-0.5 < d < 0.5$,

and $u_t$ is a stationary and Ergodic process with a bounded and positively, valued spectrum
at all frequencies. One important class of process occurs when $u_t$ is I(0) and is covariance
stationary. For $0<d<0.5$, the process is long memory in the sense of the condition (2.3.1),
its autocorrelations are all positive and decay at a hyperbolic rates. For $-0.5 < d < 0$, the
sum of absolute values of the processes autocorrelations tends to a constant, so that it has
short memory according to definition (2.3.1). In this situation the ARFIMA (0,d,0)
process is said to be ‘antipersistent’ or to have ‘intermediate memory’, and all its
autocorrelations, excluding lag zero, are negative and decay hyperbolically to zero.

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

\[
S_T = \sum_{t=1}^{T} y_t
\]  
\[(2.3.3)\]

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

\[
\sigma^2 = \lim_{T \to \infty} E \left( T^{-1} S_T^2 \right)
\]  
\[(2.3.4)\]

exists and is non zero, and

\[
[1 | \sigma T^{1/2}] S_{[rT]} \Rightarrow B(r) \text{ for all } r \in [0,1],
\]  
\[(2.3.5)\]

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

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

\[
\{ \gamma_k \cong S(k) k^{2H-2} \}
\]  
\[(2.3.6)\]
Where \( \approx \) denotes approximately equality for large \( k \) and \( S(k) \) is any slowly varying function at infinity and is described in detail by Resnick (1987). Helson and Sarason (1967) show that any process with \( H > 0 \) and autocovariance function given by (2.3.6) violate the strong mixing condition, and hence is long memory or long range dependent.

Taqqu (1975) studies the weak convergence of a linear linear combination of a long memory type process, where the weights are functions of Hermite polynomials. Specifically, the results are for a stochastic process, \( \sum_{t=1,N_p} H_m(y_t) \), where \( y_t \) is gaussian with zero mean and an auto covariance function obeying (6), \( 0 \leq p \leq 1 \), and \( H_m \) is the \( m \)-th Hermite polynomial. For the case \( H < [1 - (1/2m)] \) then an appropriately normalized version of \( \sum_{t=1,N_p} H_m(y_t) \), will converge to Brownian motion. However when \( [1-(1/2m)] < H < 1 \), the limit depends on \( m \), is non – Gaussian for \( m \geq 2 \), and when \( m = 2 \), the limit is the Rosenblatt process. Fox and Taqqu (1985) have provided further results for the quadratic form,

\[
\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j} H_m(y_i) H_n(y_j),
\]

where \( a_{i,j} \) are finite constants, \( H_m(.) \) again denote the \( m \)-th Hermite polynomial, and \( y_t \) is the same long memory process as before. The constants in the quadratic form have to decay at sufficient speed to offset of the long range dependencies in \( y_t \).

### 2.3.2 Fractional Brownian motion

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

$$B_H(r) = \left[1 | \Gamma(H+\frac{1}{2}) \right] \int_0^r (r-x)H^{-\frac{1}{2}} \, dB(x) \text{ for } r \in (0,1)$$

(2.3.7)

where $\Gamma(.)$ is the gamma function, $B(x)$ is the regular Brownian motion with unit variance, and $H$ is the Hurst coefficient, originally due Hurst (1951). When $H=\frac{1}{2}$, $B_H(r)$ reduces to regular Brownian motion, $B(r)$. The auto covariance function of Fractional Brownian motion is given by

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

and

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

(2.3.8)

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

### 2.3.3 Fractional white noise.

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

\[(1-L)^d (y_t - \mu) = \xi_t, \quad (2.3.9)\]

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

\[y_t = \sum_{k=0}^{\infty} \pi_k y_{t+k} + \xi_t, \quad (2.3.10)\]

where the infinite-order autoregressive representation weights \(\pi_k\) are obtained from the binomial expansion.

\[(1-L)^d = \{1-dL+d(d-1)L^2/2!-d(d-1)(d-2)L^3/3! + \ldots\}, \]

for any real \(d > -1\). The expansion can also be represented in terms of the hyper geometric function,

\[(1-L)^d = \sum_{k=0}^{\infty} \Gamma(k-d) L^k / \Gamma(k+1) \Gamma(-d) = F(-d, 1,1;L), \quad (2.3.11)\]

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

\[F(a,b;c;z) = \Gamma(c)/[\Gamma(a) \Gamma(b)] \]

\[x \sum_{i=1}^{\infty} z^i \Gamma(a+i) \Gamma(b+i)/[\Gamma(c+i) \Gamma(i+1)] \quad (2.3.12)\]
the typical autoregressive coefficient at lag $k$, given by $\pi_k$, is

$$\pi_k = \{d(d-1)(d-2)\ldots(d-k+1)(-1)^k\}/k! \quad (2.3.13)$$

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

and since

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

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

$$\pi_k = \Gamma(k-d)/\{ \Gamma(-d) \Gamma(k+1) \}. \quad (2.3.14)$$

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

$$y_t = \sum_{k=0}^{\infty} \Psi_k e_{t-k} \quad (2.3.15)$$

$$= (1-L)^{-d}e_t$$

$$= \{1+dL+d(d+1)L^2/2! + d(d+1)(d+2)L^3/3! + \ldots \} e_t$$

Since

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

it follows that

$$\Psi_k = \Gamma(k+d)/\{ \Gamma(d) \Gamma(k+1) \}. \quad (2.3.16)$$
The cumulative impulse response is the total impact of a unit innovation and is given by

\[ \Psi(1) = \sum_{j=0}^{\infty} \Psi_j, \]  
and the spectral density at the zero frequency is \( f(0) = \Psi(1)^2 \sigma^2 \) for \( d < 0 \) and \( f(0) = \infty \) for \( d > 0 \).

Brockwell and Davis (1987) show that \( y_t \) is convergent in mean square through its spectral representation. Also, since \( \sum_{j=0}^{\infty} \Psi_j^2 < \infty \), the fractional white noise process is mean square summable and stationary for \( d < 0.5 \). When \( d = 0.5 \), the ARFIMA \((0,0.5,0)\) process is a discrete time version of ‘1/f’ noise and is just non stationary, since \( \Psi_k \approx k^{-1/2} \), and hence \( \sum \Psi_k^2 \) just fails to converge. Odaki (1993) discusses invertibility in the sense that the MSE of the one-step-ahead linear predictor from the finite-order AR(p) converges to the innovation variance. ‘Invertibility’ in this sense was originally discussed by Granger and Andersen (1978), and

Odaki (1993) shows that \( d > -1 \) is a sufficient condition for the ARFIMA process.
### Table 2.3.1

#### 2.3.4 Properties of fractional white noise

<table>
<thead>
<tr>
<th>Infinite MA representation coefficients</th>
<th>Asymptotic approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi_k = \Gamma(k+d)/{ \Gamma(d) , \Gamma(k+1) }$</td>
<td>$\Psi_k \approx (\Gamma(d))^{-1} k^{d-1}$</td>
</tr>
<tr>
<td>$\Psi_k = \Psi_{k-1} , {(k-1+d)/k}$</td>
<td></td>
</tr>
<tr>
<td>$\Psi_k = \Pi_{0 \leq j \leq k} , [(j-1+d)/j]$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infinite AR representation coefficients</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_k = \Gamma(k-d)/{ \Gamma(-d) , \Gamma(k+1) }$</td>
<td>$\pi_k \approx (\Gamma(-d))^{-1} k^{-d-1}$</td>
</tr>
<tr>
<td>$\pi_k = \pi_{k-1} , {(k-1-d)/k}$</td>
<td></td>
</tr>
<tr>
<td>$\pi_k = \Pi_{0 \leq j \leq k} , [(j-1-d)/j]$</td>
<td></td>
</tr>
</tbody>
</table>

#### Autocovariances

| $\gamma_0 = \sigma^2 \Gamma(1-2d)/\{ \Gamma^2(1-d) \}$ |                        |
| $\gamma_k = \{ \sigma^2 \Gamma(k+d) \, \Gamma(1-2d) \}/\{ \Gamma(k+1+d) \, \Gamma(1-d) \, \Gamma(d) \}$ |                        |
| $\gamma_k = \{(\sigma^2/2\pi)\sin (nd) \, \{ \Gamma(k+d) \, \Gamma(1-2d) \}\}/\{ \Gamma(k+1+d) \}$ |                        |
| $\gamma_k = (-1)^k \, \Gamma(1-2d)/\{ \Gamma(1-k-d) \, \Gamma(k+1-d) \}$ |                        |

#### Autocorrelations

| $\rho_1 = d/(1-d), \, \rho_2 = d(1+d)/(1-2d)$, .... |                        |
| $\rho_k = \Pi_{0 < i \leq k} \, [(i-1+d)/(i-d)]$ | $\rho_k = \{ \Gamma(1-d) / \Gamma(d) \} k^{2d-1}$ |
| $\rho_h = \{ \Gamma(k+d) \, \Gamma(1-d) \}/\{ \Gamma(k-d+1) \, \Gamma(d) \}$ |                        |

#### Partial autocorrelations

| $\phi_{kh} = d/(k-d)$ for k=1,2 |                        |

#### Power spectrum

| $f(\omega) = (\sigma^2/2\pi) \, [1-e^{-i\omega t}]^{-2d}$ and at low frequencies $f(\omega) \approx c \omega^{-2d}$ | $f(0) < \infty$ if $d \leq 0$ |
| $f(\omega) = (\sigma^2/2\pi) \, [2\sin(\omega/2)]^{-2d}$ |                        |

The most important properties of the fractional white noise process are summarized in the above Table 2.3.1. and were all derived by Granger (1980), Granger and Joyeux (1980),
and Hosking (1981). Of particular interest are the long-run properties of some of the characteristics of the process. On using Stirling’s approximation for large k that 
\[
\frac{\Gamma(k+a)}{\Gamma(k+b)} \approx k^{a-b},
\]

it can be established that 
\[
\Psi_k \approx c_1 k^{d-1}, \ \pi_k \approx c_2 k^{d-1}, \ \text{and} \ \rho_k \approx c_3 k^{2d-1},
\]

where the \( c_i \) are constants. Hence the impulse response weights, infinite autoregressive coefficients, and autocorrelation coefficients all exhibit slow hyperbolic decay for large k.

Also, the power spectrum \( f(\omega) \) of the process is closely approximated at low frequencies, as \( \omega \to 0 \) by \( f(\omega) = \omega^{-2d} \) compared with \( f(\omega) \approx \omega^2 \) for a I(1), unit root process. Hence fractional white noise is consistent with the ‘typical spectral shape’ of many economic time series originally noted by Granger (1966), and the ARFIMA model to be discussed in the next section can be useful in representing spectral density functions at low frequencies. This is in contrast to differenced series which have a power spectrum that is close to being zero at low frequencies. Also, for any constant \( c \) and low frequency \( \omega \), \( f(\omega) = |c|^{2d} f(c\omega) \), so that the process is self-similar at low frequencies with the properties of \( y_t \) remaining invariant to the time interval.

A further property discussed by Sowell (1990) concerns the behavior of the contiguous sum \( S_T \) in (2.3.3), when \( y_t \) is fractional white noise as in (2.3.9). Then Sowell (1990) shows that

\[
\text{var}(S_T) = \sigma^2 \Gamma(1-2d) \{1+2d\} \Gamma(1+d) \Gamma(1-d) \}^{-1} c, \quad (2.3.17)
\]

Where

\[
c = \left[ \frac{\Gamma(1+d+T)}{\Gamma(T-d)} \Gamma(1+d)/ \Gamma(-d) \right]
\]
and

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

Hence,

\[ \text{var}(S_T) = O(T^{2d+1}), \quad (2.3.18) \]

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

\[ R(k) = k \sigma_1^2/\sigma_k^2, \quad (2.3.19) \]

for positive integer valued \( k \), and where \( \sigma_1^2 = \text{var}(y_{t-k}) \) and \( \sigma_k^2 \sim O(k^{2d-1}) \) for an I(d) process. If \( d < \frac{1}{2} \), the variance time function becomes flat; for \( \frac{1}{2} < d < 1 \), then \( R(k) \) grows at a decreasing rate, and for \( 1 < d < 3/2 \), then \( R(k) \) will grow at an increasing rate. Diebold (1989) tabulates the fractiles of \( R(k) \).

### 2.4 The ARFIMA Process

An important and more flexible class of process in discrete time has been introduced by Granger and Joyeux (1980), Granger (1980, 1981), and Hosking (1981), and is the ARFIMA \((p,d,q)\) model.
\( \phi(L)(1-L)^d(y_{t-\mu}) = 0 (L) \varepsilon_t, \quad (2.4.1) \)

Where \( d \) denotes the fractional differencing parameter, all the roots of \( \Theta(L) \) and \( \Theta(L) \) lie outside the unit circle, and \( \varepsilon_t \), is white noise. The \( y_t \) process defined by (2.4.1) and for \( d \neq 0 \) is then said to be I(d). The Wold decomposition and autocorrelation coefficients will all exhibit a very slow rate of hyperbolic decay. For \(-0.5 < d < 0.5\), the process is covariance stationary, while \( d < 1 \) implies mean reversion, for an I(d) process, the spectral density is such that \( f(0) = 0 \) for \( d < 0 \) and \( f(0) = \infty \) for \( d > 0 \). For small frequencies, \( \omega \), an approximation for \( d > 0 \) is given by \( f(\omega) \approx \omega^{-2d} \), while the process has infinite variance for \( d > 0.5 \). In particular, for the fractional process, Sowell (1986, 1992a) shows that

\[
\gamma_k = \sigma^2 \sum_{j=1}^{p} \xi_j \sum_{n=0}^{q} \sum_{m=0}^{q} \theta_n \theta_m C(d, d, p+n-m-k, \lambda_j) \quad (2.4.2)
\]

Where \( \lambda_j \) is the jth root of the AR polynomial lag operator, \( \phi(L) \), and

\[
\xi_j = [\lambda_j \prod_{i=1,p} (1 - p_i \rho_i) \prod_{k=1,p,k \neq i} (p_i - p_k)]^{-1} ,
\]

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

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

Parametric expressions for the infinite autoregressive and infinite moving average representation weights for the general ARFIMA(p,d,q) process are relatively complicated functions of the hyper geometric function. Chung (1994a) provides some alternative methods for the calculation of these autocorrelations. In particular,
\[ \gamma_k = \sum_{j=-q}^{q} a_j \sum_{n=1}^{p} \theta_n C(k-p-j; \Psi_n). \]

where

\[ C(k-p-j; \Psi_k) = \psi_k^{2p} \sum_{m=0}^{\infty} \psi_k^m \gamma_{k-p-j-m} + \sum_{n=1}^{\infty} \psi_k^n \gamma_{k-p-j+n} \]

\[ \gamma_k \] is the autocovariance at lag k of an ARFIMA (0,d,0) process,

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

\[ \psi_j = \sum_{i=0}^{\infty} \theta_0 \theta_i j. \]

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

\[ \rho_k \approx c k^{2d-1}, \quad (2.4.3) \]

where \( c > 0 \). Also, the spectral density function is

\[ f(\omega) = (\sigma^2/2\pi) |\theta(\omega)|^2 |\Phi(\omega)|^2 |1-e^{-i\omega}|^{-2d} \]

\[ = (\sigma^2/2\pi) |\theta(\omega)|^2 |\Phi(\omega)|^2 [2|1-\cos(\omega)|]^{-2d} \quad (2.4.4) \]

and for low frequencies as \( \omega \to 0 \),

\[ f(\omega) \approx (\sigma^2/2\pi)|\theta(1)/\Phi(1)|^2 \omega^{-2d} \quad (2.4.5) \]
It is often important to derive the impulse response weights from the ARFIMA \((p,d,q)\) process in (2.4.1). Following Campbell and Mankiw (1987), the impulse response weights are defined by first differencing \(y_t\) in (2.4.1), to obtain

\[(1 - L) y_t = A(L) \varepsilon_t,\]

where

\[A(L) = (1-L)^{1-d} \varnothing (L)^{-1} \theta(L).\]  

(2.4.6)

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

\[A(L) = F(d-1, 1,1;L) \varnothing(L)^{-1} \theta(L),\]

and from Gradsztein and Ryzhnic (1980, pp.1039-1042), \(F(d-1, 1,1;L)=0\) for \(d < 1\). Hence, for \(d < 1\),

\[A(1) = F(d-1, 1,1;1) \varnothing(1)^{-1} \theta(1)=0,\]  

(2.4.7)

The impact of a unit innovation at time \(t\) on the process \(y_{t+k}\) is then given as \(1+ \sum_{j=1}^{k} A_j\).

For a mean reverting process, \(A(1)=0\). For any process \(y_t \sim I(d)\) and for \(d < 1\), it follows from (2.4.7) that \(y_t\) will be mean reverting. While \(y_t\) will not be covariance stationary for \(0.5 < d < 1\), it will nevertheless still be mean reverting.

2.5 Aggregation and ARFIMA

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

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

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

$$y_{it} = \alpha_i y_{it-1} + \epsilon_{it}$$

(2.5.1)

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

$$dF(\alpha) = \frac{[2/B(p,q)]}{\Gamma(p,q)} \alpha^{p-1} (1-x^2)^{q-1}, 0 \leq \alpha \leq 1, p > 0, q > 0,$$

and shows that in the limit for large $N$,

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

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

Granger and Joyeux (1980) and Geweke and Porter-Hudak (1983) consider prediction from an ARFIMA process in (2.4.1) by using the infinite autoregressive representation

\[ y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + \alpha_t \]  

(2.6.1)

where

\[ \pi(L) = (1-L)^d \phi(L) \theta (L)^{-1} \]

They consider prediction based on a version of (2.6.1) truncated after k lags. Peiris (1987) and Peiris and Perera (1988) discuss some of the formulae for calculating predictions from the autoregressive representation. Since the ARFIMA process is not compatible with any finite-dimensional state space representation, there is no readily available solution to the truncation problem associated with using the autoregressive representation for prediction. A further currently unresolved issue concerns the effect of parameter estimation in ARFIMA processes, and the extent to which this increases prediction uncertainty. These effects may well be substantial in small samples, and are an area worthy of further investigation. Ray (1993a) considers a related issue concerning the asymptotic prediction MSE from approximating fractional white noise with a finite-order AR(p) model with estimated parameters. She finds the quality of the approximation to be very sensitive to both the order of the approximating auto regression and to the forecast horizon.
2.7 Gegenbauer and generalized ARFIMA processes

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

\[(1-2\xi L + L^2)^\lambda y_t = \varepsilon_t\quad (2.7.1)\]

The process is covariance stationary if (i) \(|\xi| < 1\) and \(0 < \lambda < 0.5\) or (ii) \(|\xi| = 1\) and \(\lambda < 0.25\). The process is invertible if (i) \(|\xi| < 1\) and \(\lambda > -0.5\) and (ii) \(|\xi| = 1\) and \(\lambda > -0.25\). When \(\xi = 1\), the process in (2.7.1) reduces to fractional white noise with \(d=2\xi\). A stationary Gegenbauer process exhibits long memory in the form a long memory harmonic behavior in its autocorrelation function. In particular, the form of the autocorrelations is dependent on the region of the parameter space. The most interesting regions is for \(|\xi| < 1\) and \(0 < \lambda < 0.5\) then

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

\[\rho_k \approx C'(\lambda)\text{ is known}\]

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

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

(2.7.3)

when \(\xi = 1\) and \(\lambda = d/2\), the Gegenbauer process in (2.7.1) reduces to the fractional white noise or ARFIMA (0,d,0) process. Hence for \(\xi = 1\) and \(0 < \lambda < 0.25\).
\[ \rho_k = \frac{\Gamma(1 - 2\lambda) \Gamma(k+2\lambda)}{\Gamma(2\lambda) \Gamma(k-2\lambda+1)}, \]

while for \( k \to \infty \),

\[ \rho_k \approx k^{4\lambda-1} \]

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

\[ \rho_k = (-1)^k \frac{\Gamma(1 - 2\lambda, k+2\lambda)}{\Gamma(2\lambda) \Gamma(k-2\lambda+1)}, \]

And as \( k \to \infty \),

\[ \rho_k \approx (-1)^k k^{4\lambda-1} \]

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

\[ \phi(L) (1-2\xi L + L^2)^\lambda (y_t-\mu) = \theta(L) \epsilon_t, \quad (2.7.4) \]

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

\[ f(\omega) = (\sigma^2 / 2\pi) |\theta(e^{-i\omega})|^2 |\phi(e^{-i\omega})|^{-\lambda} [2|\cos(v) - \cos(\omega)|]^{-2d} \quad (2.7.5) \]
As \( \omega \to v \), then the spectral density function in (2.7.5) becomes \( f(\omega) \to |\omega^2 - v^2|^{-2d} \), so that the spectral density function is unbounded as \( \omega \to v \). The form of the autocovariance function and the impulse response weights are known for lag \( k \to \infty \) to be

\[ \gamma_k \approx \cos(kv)k^{2d-1} \]  

(2.7.6)

and

\[ \Psi_k \approx \cos(kv+d(v-0.5 \pi))k^{d-1} \]  

(2.7.7)

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

\[ (1-L^s)^d y_t = \varepsilon_t, \]  

(2.7.8)

Where ‘s’ is the seasonal period, and analogously to (2.3.15), the process will have an infinite moving average representation given by

\[ y_t = \Psi(L)\varepsilon_t \]

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

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

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

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

\[ \Psi_k \approx ck^{d-1} \]  

(2.7.9)
The autoregressive function at lag k is given by

\[ \gamma_k = \sigma^2 \cos(k\pi/s) \Gamma(1-2d)/\left[ \Gamma(1-d+k|s) \Gamma(1-d-k/s) \right] \]  

(2.7.10)

The rates of decay of the autoregressive representation weights, the autocorrelations and the infinite moving average representation coefficients exactly coincide with those of the fractional white noise process in Table 2.3.1. A more general seasonal ARFIMA, or ARFISMA, model is given by

\[ \phi(L)(1-L^s)^d y_t = \theta(L)\varepsilon_t \]  

(2.7.11)

Closed form expressions for the autocorrelation function of the above process are currently unavailable. However, the spectrum of the ARFISMA model in (2.7.11) is given by

\[ f(\omega) = \left( \sigma^2 / 2\pi \right) \left| \theta(e^{-iw}) \right|^2 \left| \phi(e^{-iw}) \right|^2 \left\{ 2[1 - \cos(s\omega)] \right\}^{-2d}. \]  

(2.7.12)

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

### 2.8 Fractional cointegration

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

\[
H(L) (1-L)^d y_t = -\gamma [(1-(1-L)^b)(1-L)^{d-b}z_t + C(L) \varepsilon_t] \tag{2.8.1}
\]

Where \( H(0) = 1 \) and \( C(1) < \infty \).

2.9 Estimation and testing

2.9.1 The rescaled range statistic

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

\[
R_T = \max_{0 \leq j \leq T} \{ \sum_{j=1}^{T} (y_i-j\bar{y}) \} - \min_{0 \leq j \leq T} \{ \sum_{j=1}^{T} (y_i-j\bar{y}) \} \tag{2.9.1}
\]

Where \( R \) is the range, \( S_T \) is the sample standard deviation, and \( \bar{y} \) is the sample mean,

\[
S_T = \{(1/T) \Sigma (y_i-\bar{y})^2\}^{\frac{1}{2}}. \tag{2.9.2}
\]

Hurst (1951), Mandelbrot and AWallis (1968), Mandelbrot and Taqque (1979), Taqque (1975, 1977), and Lo (1991) showed that

\[
\lim_{T \to \infty} \{T^{-H} (R_T/S_T)\} = \text{constant} \tag{2.9.3}
\]
The idea of R/S analysis introduced by Hurst (1951) is to very informally write the above as

$$\log [E(R_t/S_t)] \approx \text{constant} + H[\log(T)],$$

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

$$Q_t = R_t/\sigma_T(q) \quad (2.9.4)$$
Where

\[ \sigma^2(q) = c_0 + 2 \sum_{j=1}^{q} w_j(q)c_j, \]  

(2.9.5)

c_j is the j^{th}-order sample autocovariance of \( y_t \) and \( w_j(q) \) are the Bartlett window weights of \( w_j(q) = 1-[j/(q+1)] \) for \( q < T \).

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

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

The distribution is positively skewed, and Lo (1991) tabulates fractiles of the distribution and shows the modified rescaled range test to be consistent against a fairly general class of long range dependent stationary Gaussian alternatives. However, a major practical difficulty concerns the choice of \( q \) and how to distinguish between short range dependencies and long range dependencies.

2.9.2 Unit root tests in the presence of \( I(d) \)

Sowell (1990) considers the limiting distribution of the OLS coefficient estimate in an AR(1) model when the true data generating process is \( I(1+d) \). When \( d=0 \), the estimate of \( \varnothing \) reduces to the well-known result derived by Phillips (1987), of
\[ T(\bar{\phi}-1) \Rightarrow (\frac{1}{2}) \left\{ B(1)^2 - 1 \right\} \left[ \int_0^1 B(t)^2 \, dt \right]. \quad (2.9.6) \]

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

It is well-known, however, that unit root tests are consistent against I(d) alternatives. A related study by Diebold and Rudebusch (1991b) evaluates the power performance by simulation of the Dickey-Fuller unit root test when the true data generating process is fractionally integrated white noise and AR(1) processes and for sample sizes of \( T=50, 100, 250 \). Not surprisingly, the power of the Dickey-Fuller tests grows more slowly with divergence of \( d \) from one that with the divergence of the AR parameter \( \bar{\phi} \) from one. Hence the Dickey-Fuller test performs relatively poorly in distinguishing between the I(1) null hypothesis and the I(d) alternative. A related study by Hassler and Wolters (1994) finds the Phillips and Perron unit root test to perform similarly to that of the Augmented Dickey-Fuller test; and with a nonstationary value of \( d=0.75 \) generating the fractional white noise, the rejection frequencies of the unit root hypothesis are about 50% when \( T=100 \) and about 70% with \( T=250 \).
Lee and Schmidt (1996) consider the performance of the KPSS test of Kwiatkowski, Phillips, Schmidt, and Shin (1992) which was originally designed to test an I(0) null hypothesis versus an I(1) alternative. The KPSS test involves taking the residuals $e_t$ from a regression of $y_t$ on an intercept and time trend and forming the partial sum $S_t$ as in (2.3.3) of the residuals and to compute the same long run variance formula $\sigma_T^2(q)$ in (2.9.5) as by Lo (1991). The KPSS test for stationary is then

$$\eta_t = T^{-2} \sum S_t^2 / \sigma_T^2(q), \quad (2.9.7)$$

and the KPSS test $\eta_t$ is also based on (48) except that the residuals are derived from a regression on an intercept only. Lee and Schmidt (1996) show that the two KPSS tests are both consistent against an I(d) alternative and that the KPSS tests can be used to distinguish short memory from long memory stationary processes. Lee and Schmidt (1996) show that under the I(d) alternative hypothesis the KPSS test statistics converge to functions of fractional Brownian motions which are relatively natural extensions of the second level Brownian bridges previously defined by MacNeill (1978) and Schmidt and Phillips (1992). Lee and Schmidt (1996) also include some Monte Carlo evidence and conclude that the KPSS test has power properties similar to the adjusted rescaled range statistic of Lo (1991) in distinguishing I(0) from I(d) behavior. Robinson (1991) derives the Lagrange Multiplier (LM) test for fractional white noise in the disturbances of a linear regression under the standard assumptions. In particular, Robinson (1991) shows that the LM test for $H_0: d=0$ versus $H_1: y_t = \beta' x_t + \epsilon_t, (1-L)^d \epsilon_t = \epsilon_t$ with $0 \leq d \leq 0.5$, is given by the statistic
\[ LM_1 = T \left\{ \sum_{j=1}^{T-1} j^{-1} r_j \right\}^2 / (\pi^2/6) \]  

(2.9.8)

Where \( r_j \) is the \( j \)th - order sample autocorrelation coefficient of the OLS residuals. Under the null, \( LM_1 \) will have \( n \) asymptotic chi-squared distribution with one degree of freedom. Interestingly, Robinson (1991) also shows that to a first-order approximation the same test statistic results from the alternative hypothesis that the process is self-similar, with autocovariance function given by (2.3.11).

Wu (1992) considers the related issue of testing the unit root hypothesis versus the one-sided alternatives of \( d < 1 \) or \( d > 1 \). The tests are modified one-sided locally best invariant (LBI) alternatives based on the technique developed by King and Hillier (1985). Wu (1992) shows that the LBI test for \( H_0: d = 0 \) versus \( H_1: d > 0 \) in the fractional white noise model in (2.3.9) is given by

\[ LM_2 = 2 \sum_{j=1}^{T-1} j^{-1} r_j \]

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

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

Taqqu (1975) and Yajima (1988) consider estimation of regression parameters in the presence of disturbances exhibiting long memory. A special case arises in the estimation of the population mean of a long memory process with autocovariance function, $\gamma_k \approx ck^{2d-1}$, as given by (2.3.6), (2.3.8), or (2.3.12), Taqqu (1975) derives the well-known result concerning the properties of the OLS estimate of the mean or intercept parameter $\mu$ in the model $y_t = \mu + \nu_t$, where $\nu_t$ has an autocovariance function of $\gamma_k \approx ck^{2d-1}$. Taqqu (1975) shows that the sample mean coverages at a rate of $T^{1/2-d}$ to an unspecified limited distribution. Furthermore, the estimator of the uncorrected sample standard deviation,

$$S_T = \left\{ T^d \sum (y_t - \bar{y})^2 \right\}^{1/2}$$  \hspace{1cm} (2.9.9)

and on appropriately normalizing,

$$cT^{-d} \{ S_T - E(S_T) \} \Rightarrow R,$$  \hspace{1cm} (2.9.10)

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

$$T^{1-2d} \text{var}(\bar{y}) = c_2 / [d(2d-1)],$$  \hspace{1cm} (2.9.11)

Where $c_2$ is another constant. Also,
\[ c T^{1-2d} (\bar{y} - E(\bar{y}))^2 \Rightarrow [d(2d-1)]^{-1} \chi^2_d \]  

(2.9.12)

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

\[ y_t = \beta' x_t + u_t, \]

where \( u_t \) is a long memory process and \( x_t \) contains polynomial functions of time. The above articles find expressions for the asymptotic efficiency loss from using OLS, rather than GLS. As in the case of the regression with just an intercept, the loss of efficiency associated with OLS is not necessarily severe. Robinson (1990) extends this investigation to the case where \( x_t \) also contains stochastic regressors. The resulting parameter estimates appear to converge at different rates and Robinson (1990) notes that in certain cases, singular limiting distributions may result. It should also be noted that Carlin, Dempster, and Jonas (1985) and Carlin and Dempster (1989) have suggested a Bayesian estimator in the context of long memory models and the Hurst coefficient.
Cheung and Lai (1993) have considered testing for the fractional cointegration of two time series, $y_t$ and $x_t$, which are both I(d) but are fractionally cointegrated, i.e., CI(d,b). Hence $\varepsilon_t = y_t - \beta x_t$ and $\varepsilon_t$ is I(d-b), where $(d-b) > 0$. In the case of $(d-b) > \frac{1}{2}$, Cheung and Lai (1993) show that the OLS estimator of $\beta$ converges in probability to zero for all $\delta > 0$, such that

$$T^{1/2} (\hat{\beta} - \beta) \sim N(0, V_k),$$

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

### 2.9.4 Distribution of sample autocorrelations from an I(d) process

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

$$T^{1/2} (f_k - \rho_k) \sim N(0, V_k),$$
and $V_K$ is derived from the asymptotic covariance matrix, using the usual formula of Bartlett.

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

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

and

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

Where ‘D’ is a non-standard distribution. Hence the rate of convergence is slower than the conventional rate in this range of values of $d$. Hosking also shows that

$$T^{1/2-d}(r_k - \rho_k) (1 - \rho_k)^{-1} - (r_j - \rho_j) (1 - \rho_j)^{-1}$$

(2.9.13)

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

### 2.9.5 Semiparametric estimation of $d$ in the frequency domain

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

\[
f(\omega)_y = |1 - e^{-i\omega t}|^{-2d} \ f(\omega)_u,
\]  

(2.9.14)

where \(f(\omega)_y\) and \(f(\omega)_u\) are the spectral densities of \(y_t\) and \(u_t\) respectively. Then (2.9.14) can be expressed as

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

\[
\log \{f(\omega)_y\} = \log\{f_u(0)\} - d \log\{4 \sin^2(\omega_j/2)\} + \log[f_u(\omega_j)/f_u(0)].
\]  

(2.9.15)

GPH suggest estimating ‘d’ from a regression based on (2.9.15) using spectral ordinates \(\omega_1, \omega_2, \ldots, \omega_m\), from the periodogram of \(y_t\), that is \(I_y(\omega_i)\). Hence, for \(j=1,2,\ldots,m\),

\[
\log\{I_y(\omega_j)\} = a + b \log \{4 \sin^2(\omega_j/2)\} + v_j,
\]  

(2.9.16)

where

\[
v_j = \log [f_u(\omega_j) / f_u(0)]
\]  

(2.9.17)

and \(v_j\) is assumed to be i.i.d. with zero mean and variance \(\pi^2/6\). When \(u_t\) is white noise, \(\varepsilon_t\), then the regression (2.9.16) should provide a good estimate of ‘d’. When \(u_t\) is autocorrelated, GPH show that (2.9.16) holds approximately for frequencies in the neighborhood of zero. If this neighborhood shrinks at an appropriate rate with sample size, then the GPH procedure should realize a consistent estimator of ‘d’. If the number of ordinates \(m\) is chosen such that \(m = g(T)\),
where \( g(T) \) is such that \( \lim_{T \to \infty} g(T) = \infty \), \( \lim_{T \to g} \{g(T)/T\} = 0 \), \( \lim_{T \to g}\{(\log(T)^2/g(T))=0\} \), then the OLS estimator of \( d \) in (2.9.16) will have the limiting distribution.

\[
\left( \hat{d}_{GPH} - d \right) / \left\{ \text{var} \, \hat{d}_{GPH} - d \right\}^{1/2} \Rightarrow N(0,1).
\]

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

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

While the GPH estimator is simple to apply and is potentially robust to non-normality, the behavior of \( (\hat{d}_{GPH}) \) in the presence of substantial autocorrelation of \( u_t \) reduces its potential attractiveness. In particular, Agiakloglou, Newbold, and Wohar (1992) show it possesses ‘serious bias’ and is very inefficient when \( u_t \) is AR(1) or MA(1) and the AR or MA parameter is quite large. Also, if an investigator wishes to obtain...
estimates of short run ARMA parameters as well as d, then filtering the original series by the operator, \((1-L)^d\) where \(d\) is replaced with \(\hat{d}_{\text{GPH}}\), and estimating the ARMA parameters from the filtered series will provide two-step estimates with a currently unknown sampling distribution. Most applications of this procedure, such as Diebold and Rudebusch (1989, 1991b), typically assume normality of the filtered series and use approximate MLE to obtain estimates of the ARMA parameters. Even if a quasi-MLE is used in this situation, appropriate inference is likely to remain a difficulty. Hassler (1994) has carried out a simulation study of a variant of the GPH procedure applied to the seasonal ARFISMA process of Porter-Hudak (1990) in (2.9.3). The results generally indicate deficiencies with the semiparametric regression. Hurvich and Ray (1995) have consolidated the bias of the GPH estimator in the case when the true data generating process is a nonstationary ARFIMA process with a value of \(d > 0.5\).

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

\[
F(\omega) = \int f(\lambda) \ d \lambda,
\]

Where the averaging is over the neighborhood \(\omega \in (0, \lambda)\). Robinson (1992) shows that \(F(\omega)\) converges in probability to one for a sequence \(\lambda\) which tends to zero more slowly than \(1/T\), as \(T \to \infty\). For any slowly varying function \(\equiv (\cdot)\), then
\[
F(q\omega)/F(\omega) \approx q^{2(H-1)} \{ \equiv (1/qw)'/\equiv (1/\omega) \}, \tag{2.9.18}
\]

\[
\approx q^{2(H-1)} \tag{2.9.19}
\]

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

\[
\hat{H}_q = 1-\{2\log(q)\}^{-1} \log \left( F((q\omega_m)/F(\omega_m)) \right), \tag{2.9.20}
\]

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

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

### 2.9.6 Semiparametric estimation of $d$ in the time domain

An alternative to periodogram based estimation is to directly use the sample autocorrelations. Robinson (1990) considers such an estimate based upon high lags of $y_k$, but notes that such a procedure has the disadvantage of being based on the requirements that the $y_k$ are always eventually positive. An alternative autocovariances, has been studied by Tieslau, Schmidt, and Baillie (1995), henceforth TSB. Their estimator uses blocks of $n$ sample autocovariances, $\gamma_k, \gamma_{k+1}, \ldots, \gamma_{k+m}$. From using the results of Hosking (1984), who derives the asymptotic distribution of sample autocovariances from fractional white noise, TSB evaluate the asymptotic efficiency for their MDE. As noted by Hosking (1984) and formally proved by Dahlhaus (1988, 1989), the unusual behavior of the score vector makes the rate of convergence to the limiting distribution dependent on the value of ‘$d$’. For $-0.5 < d < 0.25$, the MDS studied by TSB converges to a limiting normal distribution at the conventional $T^{1/2}$ rate, while for $d = 0.25$, the MDE still converges to a normal distribution, but at a rate of $\{T/\log(T)\}^{1/2}$. For $0.25 < d < 0.50$, the MDE converges at rate of $T^{1/2-d}$ to a nonstandard distribution. TSB then evaluate the theoretical asymptotic efficiency of the MDE for various parameter values ‘$d$’ in the range $-0.50 < d < 0.25$ and for different choices of $k$ and $n$. The efficiency loss for increasing $k$ is found to be quite large and the estimator is not generally very promising.
The results of Hosking (1984) in (2.9.18) suggest that it may be possible to obtain a semiparametric estimate of \( d \) in the time domain that converges at the standard rate, for the case where \( 0.25 \leq d < 0.50 \). However, some simulation evidence indicates that the function form of (2.9.18) may not be very helpful and that an estimator based upon differenced autocorrelations will be quite inefficient.

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

\[
\gamma'_k \approx \left[\Theta(1)/\Phi(1)\right] \gamma_k,
\]

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

### 2.9.7 Maximum likelihood estimation

Several authors have considered joint estimation of the parameters in the ARFIMA\((p,d,q)\) model in (2.4.6) under the assumption of normality. The \((p+q+3)\) dimensional vector of parameters is \( \lambda' = (\mu \beta') \), where

\[
\beta' = (d \varnothing \ldots \varnothing_p \theta_1 \ldots \theta_q \sigma^2)
\]
Li and McLeod (1986) consider the asymptotic properties of the MLE in the case of the intercept $\mu$ being either known or zero. They assert that

$$T^{1/2} \{ (\hat{\beta} - \beta) \} \to N[0, \tau \lim_{T \to \infty} \{I(\beta)/T\}^{-1}], \quad (2.9.21)$$

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

$$I(\beta) = \begin{pmatrix} I_{p,q} & J \\ J' & \pi^2/6 \end{pmatrix}, \quad (2.9.22)$$

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

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

Where

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

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

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

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

For the fractional white noise process, $(1-L)^d y_t = \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$, then (2.9.21) reduces to the well-known result that
The above is an example of the general and rather surprising property of ARFIMA models that the asymptotic variance of their parameter estimates are independent of the value $d$. This is contrast to ARMA models, where each parameter generally occurs in at least one element of the information matrix. The form of $I(\beta)$ in (2.9.22) also implies that the MLE of ‘$d$’ will generally by asymptotically correlated with the ARMA parameter estimates. While Galbraith an Galbraith (1974) and Newbold (1974) provide parametric results for the inverse of $I(\lambda)$ for the class of stationary and invertible ARMA models, no corresponding results are yet available for the ARFIMA process.

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

$$I_{jk}(\lambda) = (1/4\pi) \int_{-\pi}^{\pi} \left\{ (2\pi/\sigma^2) \{ \log f(\omega|\lambda) / \delta\lambda_j \} \right\} \times \left\{ (2\pi/\sigma^2) \{ \delta \log f(\omega|\lambda) / \delta\lambda_k \} \right\} d\omega. \quad (2.9.24)$$

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

$$D_{T}(\hat{\lambda} - \lambda) \Rightarrow N \left\{ 0, [D_T^{-1} I(\lambda) D_T^{-1}]^{-1} \right\}, \quad (2.9.25)$$

Where

$$T^{1/2}(\hat{d} - d) \Rightarrow N(0,6/\pi^2). \quad (2.9.23)$$
diag \{D_T\} - [T^{1/2-d}, T^{1/2}, \ldots, T^{1/2}] \quad (2.9.26)

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

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

\[
\mathcal{L} = -(T/2)\log (2\pi) = -(1/2)\log |\Sigma| - (1/2)Y\Sigma^{-1}Y, \quad (2.9.27)
\]

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

There are several alternative approximate MLE of the ARFIMA(p,d,q) model in (2.4.6) under normality. Whittle (1951) notes that the autocovariance matrix \( \Omega \) can be diagonalized by transforming the vector \( Y \) into the frequency domain and can approximate the log-likelihood by

\[
\mathcal{L} = \sum_{j=1}^{T-1} \log[(2\pi) f(\omega_j)] + \sum_{j=1}^{T-1} \left[ I_1(\omega_j) / f(\omega_j) \right]. \quad (2.9.28)
\]

The above approximate MLE has been used by Boes, Davis, and Guptha (1989), who concentrate out \( \sigma^2 \) to obtain
\[ \sigma^2(\lambda) = (2\pi/T) \sum_{j=1}^{T-1} \left[ I_T(\omega_j) / f(\omega_j) \right] \]

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

\[ \Sigma \{I(\omega_j)\} / f(\omega_j, 0), \quad (2.9.29) \]

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

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

\[ S = \left( \frac{1}{2} \log (\sigma^2) + \frac{1}{2} \sigma^2 \right) \sum_{t=1}^{T} \left\{ \theta(L) \theta(L)^{-1} (1-L)^d (y_t-\mu) \right\}^2 \quad (2.9.30) \]

Some results concerning the small sample performance of the CSS estimator are reported in Chung and Baillie (1993). They conclude that for the ARFIMA (0,d,0) model, with \( T=100 \) and with the mean unknown, CSS is extremely similar to Sowell’s full MLE. For the ARFIMA (p,d,q) model with unknown mean and complicated ARMA dynamics, i.e. \( p,q > 2 \), the CSS estimator can produce substantial biases in samples of 300. The estimation of the intercept \( \mu \) can substantially affect the properties of the other parameter estimates. However, the CSS estimator performs quite well for ARFIMA models with known mean parameter and \( T = 500 \). Some further simulation evidence is provided by Cheung and Diebold (1994). Interestingly enough they find that the Fox-Taqqu estimator
is preferable to Sowell’s full MLE when the mean of the process, $\mu$, is unknown. In their application of the Fox-Taqqu estimator, Cheung and Diebold (1994) essentially use a two-step estimator which replaces the unknown mean parameter $\mu$ with the sample mean of $y_i$ before estimating the other parameters with the Fox-Taqqu estimator. Cheung and Diebold (1994) find evidence that their estimator is more satisfactory than the full MLE of Sowell in the sense of bias and MSE. These differences are again brought about by the key fact that any estimator of $\mu$ converges at a slower rate than the other parameter estimates which are all $T^{1/2}$ consistent. In some applications (e.g. Sowell, 1992a) the data series are differenced and the process is estimated with $d < 0.50$. This strategy has the advantage of removing the troublesome intercept parameter, and some simulation evidence on the efficacy of the procedure is presented by Smith, and ZIN (1993). A further time domain approximate MLE is described by Haslett and Raftery (1989), who only consider $d$ in the range of $0 \leq d \leq 0.5$. While the full MLE is obviously desirable when appropriate, more complicated models can only currently be estimated by approximate MLE, which can be conveniently done by minimizing the CSS function in (2.9.30). For example, Baillie, Chung, and Tieslau (1995) estimate an ARFIMA process with a conditional variance process following a GARCH (1,1) formulation. They estimate monthly inflation series with an ARFIMA (0,3,1) – Garch (1,1) process.

Chung (1996) provides some results on the estimation of the Gegenbauer process in (2.7.6) by approximate MLE, using the CSS method. An important finding is that the MLE of $\xi$ converges at a rate of $O_p(T)$ to a function of Brownian motions. The other parameter estimates converge at the usual $T^{1/2}$ rate.
2.9.8 Long memory volatility processes

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

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

\[
\epsilon_t = \xi_t \sigma_t
\]  

(2.9.31)

where \( \xi_t \) is i.i.d with \( E(\xi_t) = 0 \) and \( \text{var}(\xi_t)=1 \). The variable \( \sigma_t^2 \) is a time-varying, positive, and measurable function of the information set at time \( t-1 \), denoted by \( \Omega_{t-1} \), and \( \sigma_t^2 \) is known as an ARCH process. The GARCH(p,q) specification of Bollerslev (1986) is defined as
\[ \sigma_t^2 = \omega + \alpha(L) \varepsilon_t^2 + \beta(L) \sigma_t^2, \quad (2.9.32) \]

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

\[ \{1 - \alpha(L) - \beta(L)\} \varepsilon_t^2 = \omega + \{1 - \beta(L)\} \nu_t, \quad (2.9.33) \]

Where \( \nu_t - \varepsilon_t^2 - \sigma_t^2 \) are the ‘innovations’ in the conditional variance process. When the polynomial \( \{1 - \alpha(L) - \beta(L)\} \) in (2.9.33) contains a unit root, then the GARCH\((p,q)\) process is a member of the integrated GARCH, or IGARCH\((p,q)\), class of models defined by

\[ \phi(L)(1-L) \varepsilon_t^2 = \omega + \{1 - \beta(L)\} \nu_t, \]

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

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

\[ \phi(L)(1-L)^d \varepsilon_t^2 = \omega + \{1 - \beta(L)\} \nu_t, \quad (2.9.34) \]
Where all the roots of $\phi(L)$ and $(1-\beta(L))$ lie outside the unit circle. Analogously to (2.9.34) the FIGARCH process can also be represented as

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

(2.9.35)

and as

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

(2.9.36)

Where

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

(2.9.37)

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

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

which is directly analogous to corresponding calculations for the mean given by (2.4.5). Then,

$$
(1-L) \varepsilon_t^2 = \omega + y(L) \nu_t
$$

(2.9.38)

and

$$
y(L) = (1-L)^{-d} \phi(L)^{-1} \{1-\beta(L)\}
$$

(2.9.39)
The impact of past shocks on the volatility process is given by the limit of the cumulative impulse response weights,

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

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

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

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

\[ \lambda_k = [\Gamma(k+d-1)/\Gamma(k) \Gamma(d)] [1-\beta] \] \( -(1-d)/k \). \( \] (2.9.40)

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

\[ \lambda_k \approx [(1-\beta) / \Gamma(d)] k^{d-1}, \] \( \] (2.9.41)

so that hyperbolic decay occurs in the response of the conditional variance to past shocks. Since \( \lambda(1)=1 \), it follows that \( E(\varepsilon_t^2) \) is undefined, and hence the second moment of the
unconditional density of $\varepsilon_t$ is infinite. The FIGARCH process is not weakly stationary, a feature it shares with the IGARCH process. Approximate maximum likelihood estimates of the parameters of the FIGARCH(p,d,q) process in (2.9.34) can be obtained by maximizing the Quasi Maximum Likelihood, which realizes $T^{1/2}$ consistent estimates of the FIGARCH parameters. Then,

$$T^{1/2}(\hat{\theta}_T - \theta_0) \Rightarrow N\{0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}\}, \quad (2.9.42)$$

Where $A(.)$ and $B(.)$ represent the Hessian and outer product gradient, respectively, and $\theta_0$ denotes the true parameter values. Simulation evidence indicates that the limiting distribution theory works well in sample sizes of 1500 and 3000. Balillie, Bollerslev, and Mikkelsen (1996) also report the effects of estimating stable GARCH processes where the true data generating process in FIGARCH. The sum of the estimated GARCH(1,1) parameters is always close to one, which implies integrated GARCH, or IGARCH, behavior and suggests that the apparent widespread IGARCH properly so frequently found in high frequency asset pricing data (see Bollerslev, Chou, and Kroner, 1992) may well be spurious, that the IGARCH process is a poor diagnostic at distinguishing between integrated, as opposed to long memory, formulations of the conditional variance process.

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

$$\log(\sigma_t^2) = \omega + \varphi(L)^d [1-\lambda(L)]g(\xi_{t-1}), \quad (2.9.43)$$

where
\[ g(\xi_t) = 0\xi_t + y[|\xi_t| - E|\xi_t|] \]  
\hspace{1cm} \text{ (2.9.44)}

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

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

\[ y_t = \xi_t \sigma_t, \]

and

\[ \sigma_t^2 = \sigma^2 \exp(h_t) \]  
\hspace{1cm} \text{ (2.9.45)}

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

\[ (1-L)^d h_t = \epsilon_t, \]  
\hspace{1cm} \text{ (2.9.46)}

where \( \epsilon_t \sim \text{NID}(0, \sigma^2) \), then (2.9.45) and (2.9.46) generate a long memory stochastic volatility process. Estimation of regular stochastic volatility models has generally been through the state space representation and use QML estimation via the Kalman filter.
Since a state space representation does not exist for long memory processes, estimation of the long memory stochastic volatility process is correspondingly difficult. Breidt, Crato, and de Lima (1993) use frequency domain approximate MLE to estimate an ARFIMA(0,d,1) model for log(y_t^2), while Harvey (1993) uses the GPH estimator to obtain an estimate of d in a fractional white noise model for log (y_t^2). The comparison of long memory ARCH and stochastic volatility models remains an interesting area for future research.