Financial Time Series

Topic 5: Fractional Integration and Long Memory Processes

Cheng-Der Fuh
Institute of Statistical Science
Academia Sinica
Spring 2003
OUTLINE

1. Decomposition of Time Series
   • Unobservable Component Model
   • Signal Extraction

2. Measures of Persistence and Trend Reversion

3. Fractional Integration and Long Memory Processes
   • *ARFIMA* models
   • Spectral Density
   • Testing for fractional differencing
   • Estimation in *ARFIMA*
Decomposition of Time Series

Unobserved component models

- If a time series is difference stationary, it can be decomposed into a **trend component** (stochastic non-stationary) and a **noise component** (stationary).
  
- Write it as

  \[
  x_t = \hat{z}_t + \hat{u}_t. \tag{1}
  \]

- Identification: Given only the observed time series \( \{x_t\} \) and its model, can we identify the two unobserved components \( z_t \) and \( u_t \)?

- Signal Extraction: How do we estimate these two unobserved components?

  The example of Muth’s (1960)

- The unobservable random walk buried in white noise.

- The trend component, \( z_t \), is a random walk

  \[
  z_t = \mu + z_{t-1} + v_t.
  \]

- The noise component, \( u_t \), is white noise and independent of \( v_t \).

- \( u_t \sim WN(0, \sigma_u^2), \quad v_t \sim WN(0, \sigma_v^2) \).

- \( \Delta x_t \) is a stationary process

  \[
  \Delta x_t = \mu + v_t + u_t - u_{t-1}. \tag{2}
  \]
• ACF of $\Delta x_t$ cuts off at lag one with coefficient

$$\rho_1 = -\frac{\sigma_u^2}{\sigma_u^2 + 2\sigma_v^2},$$

where $\sigma_u^2 + 2\sigma_v^2$ is the variance of $\Delta x_t$.

• Note that $-0.5 \leq \rho_1 \leq 0$ and the exact value depending on $\kappa = \sigma_v^2/\sigma_u^2$, the signal-to-noise variance ratio.

• When $\kappa = 0$, $z_t$ is a deterministic linear trend, while $\kappa = \infty$ dictates that $x_t$ is a pure random walk.

• We can write $\Delta x_t$ as an $MA(1)$ process

$$\Delta x_t = \mu + e_t - \theta e_{t-1},$$

where $e_t \sim WN(0, \sigma_e^2)$.

• It can be shown that

$$|\theta| < 1; \quad \theta = \left\{ (\kappa + 2) - (\kappa^2 + 4\kappa)^{1/2} \right\} / 2;$$

$$\kappa \geq 0; \quad \kappa = (1 - \theta)^2 / \theta;$$

$$\sigma_u^2 = \theta \sigma_e^2.$$

• Identifiability: $\sigma_u^2$ is the lag one autocovariance of $\Delta x_t$. $\sigma_v^2$ can be calculated from the variance of $\Delta x_t$ and $\sigma_u^2$. 
General UC Models

\[ \Delta z_t = \mu + \gamma(B)v_t, \quad u_t = \lambda(B)a_t, \]

where \( v_t \) and \( a_t \) are independent white noise sequences with finite variance \( \sigma_v^2 \) and \( \sigma_a^2 \) and where \( \gamma(B) \) and \( \lambda(B) \) are stationary polynomials having no common roots.

- Hence \( \Delta x_t = \mu + \theta(B)e_t \),
  where \( \theta(B) \) and \( \sigma_e \) can be obtained from

\[
\sigma_e^2 \frac{\theta(B)\theta(B^{-1})}{(1 - B)(1 - B^{-1})} = \sigma_v^2 \frac{\gamma(B)\gamma(B^{-1})}{(1 - B)(1 - B^{-1})} + \sigma_a^2 \lambda(B)\lambda(B^{-1}).
\]

The parameters can not be identified in general.

- Poterba and Summers (1988):

\[ u_t = \lambda u_{t-1} + a_t, \]

so that

\[ \Delta x_t = \mu + v_t + (1 - \lambda B)^{-1}(1 - B)a_t \]

or

\[ \Delta x^*_t = (1 - \lambda)\mu + (1 - \lambda B)v_t + (1 - B)a_t, \]

where \( x^*_t = (1 - \lambda B)x_t \). Thus \( \Delta x_t \) follows an \( ARMA(1, 1) \) process

\[ (1 - \lambda B)\Delta x_t = \theta_0 + (1 - \theta_1)e_t, \]

where \( \theta_0 = \mu(1 - \lambda) \). Hence

\[ \theta_1 = \frac{\{2 + (1 + \lambda)^2\kappa - (1 - \lambda)[(1 + \lambda)^2\kappa^2 + 4\kappa]^{1/2}\}}{2(1 + \lambda\kappa)}, \]

and \( \sigma_e^2 = (\lambda\sigma_v^2 + \sigma_a^2)/\theta_1 \).
Signal Extraction

- The MMSE of $z_t$ is an estimate $\hat{z}_t$ which minimizes $E(\zeta_t^2)$, where $\zeta_t$ is the estimation error $z_t - \hat{z}_t$.
- Given $\{x_t\}_{-\infty}^{\infty}$, Pierce (1979) proposes

$$\hat{z}_t = \nu_z(B)x_t = \sum_{j=-\infty}^{\infty} \nu_{zj}x_{t-j},$$

where the filter $\nu_z(B)$ is defined as

$$\nu_z(B) = \frac{\sigma_v^2\gamma(B)\gamma(B^{-1})}{\sigma_e^2\theta(B)\theta(B^{-1})}.$$

- The noise component can be estimated as

$$\hat{u}_t = x_t - \hat{z}_t = [1 - \nu_z(B)]x_t = \nu_u(B)$$

- For example, under the Muth model, we have

$$\nu_z(B) = \frac{\sigma_v^2}{\sigma_e^2} (1 - \theta B)^{-1} (1 - \theta B^{-1})^{-1}$$

$$= \frac{\sigma_v^2}{\sigma_e^2(1 - \theta^2)} \sum_{j=-\infty}^{\infty} \theta^{|j|} B^j.$$

- Note that $\sigma_v^2 = (1 - \theta)^2 \sigma_e^2$, we have

$$\hat{z}_t = \frac{(1 - \theta)^2}{1 - \theta^2} \sum_{j=-\infty}^{\infty} \theta^{|j|} x_{t-j}.$$

- For values of $\theta$ close to 1, $\hat{z}_t$, will be given by a very long moving average of future and past values of $x$. If $\theta$ is close to 0, however, $\hat{z}_t$, will be almost equal to the most recent observed value of $x$. 

• The estimation error, \( \zeta_t = z_t - \hat{z}_t \), can be written as
\[
\zeta_t = \nu_z(B)z_t - \nu_u(B)u_t.
\]
• Pierce (1979) shows that \( \zeta_t \) will be stationary if \( z_t \) and \( u_t \) are generated by process of the form (4).
• In fact, \( \zeta_t \) will follow the process
\[
\zeta_t = \theta \zeta(B)\xi_t,
\]
where \( \theta \zeta(B) = \frac{\gamma(B)\lambda(B)}{\theta(B)} \), \( \sigma^2_{\xi} = \frac{\sigma^2_a\sigma^2_v}{\sigma^2_e} \).
• For the Muth model, we thus have that \( \zeta_t \) follows the AR(1) process
\[
(1 - \theta B)\zeta_t = \xi_t
\]
and the MSE of the optimal signal extraction procedure is
\[
E(\zeta^2_t) = \frac{\sigma^2_a\sigma^2_v}{\sigma^2_e(1 - \theta^2)}.
\]
• If \( x_t \) follows the IMA(1, 1) process
\[
(1 - B)x_t = (1 - \theta B)e_t
\]
then the most general signal-plus-white-noise UC model has \( z_t \) given by
\[
(1 - B)z_t = (1 - \Theta B)v_t.
\]
• For any \( \Theta \) value in the interval \( -1 \leq \Theta \leq \theta \) there exist values of \( \sigma^2_a \) and \( \sigma^2_v \) such that decomposition is possible.
• It can be shown that setting $\Theta = -1$ minimizes the variance of both $z_t$ and $u_t$ and is known as canonical decomposition of $x_t$.

• Choosing this value implies that $\gamma(B) = 1 + B$ and we thus have

$$\hat{z}_t = \frac{\sigma^2_v(1 + B)(1 + B^{-1})}{\sigma^2_e(1 - \theta B)(1 - \theta B^{-1})}$$

and

$$(1 - \theta B)\zeta_t = (1 + B)\xi_t.$$ 

• How to estimate $z_t$ if we only have data on $\{x_t\}$ up to time $t - m$?

• Pierce (1979) proposed the following:

For $m \geq 0$,

$$\hat{z}^{(m)}_t = (1 - \theta)B^m \sum_{j=0}^{\infty} (\theta B)^j x_t.$$ 

For $m < 0$,

$$\hat{z}^{(m)}_t = \frac{1 - \theta}{\theta^m} B^m \sum_{j=0}^{\infty} (\theta B)^j x_t + \frac{1}{1 - \theta B} \sum_{j=0}^{-m-1} \theta^j B^{-j} x_t.$$ 

Example 3.5 Estimating Expected Real Rates of Interest

- What is the expected real rates of interest under the assumption of rational expectation (financial market efficiency).

- \( z_t \): unobservable expected real rate, which is assume to follow a driftless random walk.

- \( u_t \): unexpected inflation, which is a white-noise process if the market is efficient.

- \( x_t \): observed real rate will thus follow the ARIMA\((0, 1, 1)\) process.

- Such a UC model fitted to the real UK Treasury bill rate over the period 1952Q1 to 1995Q3 yields

  \[
  \Delta x_t = (1 - 0.694B)e_t, \quad \hat{\sigma}_e^2 = 7.62, \quad Q(12) = 9.1.
  \]

- Hence,

  \[
  \hat{\sigma}_v^2 = (1 - 0.694)^2 \hat{\sigma}_e^2 = 0.71 \\
  \hat{\sigma}_u^2 = 0.694 \hat{\sigma}_e^2 = 5.29.
  \]

- The variations in the expected real rate are small compared to variations in unexpected inflation.

  \[
  \kappa = \sigma_v^2 / \sigma_u^2 = 0.71 / 5.29 = 0.134.
  \]

- Using exponentially weighted moving average

  \[
  \hat{z}_t = v_Z^{(0)}(B)x_t = (1 - 0.694) \sum_{j=0}^{\infty} (0.694B)^j x_t.
  \]
• Unexpected inflation can be obtained as \( \hat{u}_t = x_t - \hat{z}_t \).

• Fig 3.8 provides plots of \( x_t, \hat{z}_t, \) and \( \hat{u}_t \) showing
  
  – The expected real rate is considerably smoother than the observed real rate due to small \( \kappa \).
  
  – In the early part of 1950s expected real rates were generally negative.
  
  – From 1956 to 1970, they were consistently positive
  
  – From the middle of 1970 and the subsequent decade, the expected real rate was always negative.
  
  – The minimum is reached in 1975Q1 after inflation peaked in the previous quarter as a consequence of the OPEC price rise, and a local minimum in 1979Q2, this being a result of VAT increase in the budget of that year.
  
  – From mid 1980 the series is again positive and remains so until the end of the sample period.
  
  – Fluctuations in unexpected inflation are fairly homogeneous except for the period from 1974 to 1982.
Figure 3.8: Interest rate decomposition
Measure of Persistence

• Suppose that $x_t$ contains a unit root. Then

$$\Delta x_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}. \quad (5)$$

• The impact of a shock in period $t$, $a_t$, on the change in $x$ in period $t + k$, $\Delta x_t$, is $\psi_k$.

• The impact of the shock on the level of $x$ in period $t + k$, $x_{t+k}$, is therefore $1 + \psi_1 + \cdots + \psi_k$.

• The ultimate impact on the level of $x$ is

$$A(1) = 1 + \psi_1 + \psi_2 + \cdots = \sum_{j=0}^{\infty} \psi_j.$$  

• The value of $A(1)$ is a measure of how persistent shocks to $x$ are.

(a) $A(1) = 0$: trend stationary series

(b) $A(1) = 1$: random walk

• Campell and Mankiw (1987) offer a measure of $A(1)$ based on approximating $A(B)$ by a ratio of finite-order polynomials.

• Because $\Delta x_t$ is stationary, it has an $ARMA(p, q)$ representation $\phi(B) \Delta x_t = \theta_0 + \theta(B) a_t$. So we have the **impulse response function** of $\Delta x_t$

$$\Delta x_t = \phi(1)^{-1} \theta_0 + \phi(B)^{-1} \theta(B) a_t.$$
• From the equality $A(B) = \phi(1)^{-1}\theta(B)$, it follows that $A(1) = \theta(1)/\phi(1)$.

• Cochrane (1988) proposes a non-parametric measure of persistence known as the variance ratio

$$V_k = \sigma_k^2/\sigma_1^2,$$

where $\sigma_k^2 = k^{-1}V(x_t - x_{t-k}) = k^{-1}V(\Delta_k x_t)$ and $\Delta_k = 1 - B^k$ being the $k$th differencing operator.

• If $x_t$ is a pure random walk with drift, then

$$V(\Delta_k x_t) = V[(x_t - x_{t-1}) + \cdots + (x_{t-k+1} - x_{t-k})]$$

$$= \sum_{j=1}^{k} V(x_{i-j+1} - x_{t-j}) = V(a_{i-j+1})$$

$$= k\sigma^2.$$

• If $x_t$ is trend stationary, the variance of the $k$th differences a constant, which is twice the unconditional variance of the series: $x_t = \beta_0 + \beta_1 t + a_t \Rightarrow$

$$V(\Delta_k x_t) = V(a_t) + V(a_{t-k}) = 2\sigma^2.$$

• When plotting the estimate of $\sigma_k^2$ as function of $k$,

(i) A random walk: the plot should be constant $\sigma^2$.

(ii) Trend stationary: the plot should decline to zero.

(iii) Partly permanent and partly temporary: the plot should settle down to the variance of the innovation to the random walk component.
• The sample estimate of $\sigma^2_k$ is as follows

$$\hat{\sigma}^2_k = \frac{T}{k(T-k)(T-k+1)} \sum_{i=k}^{T} \left[ x_t - x_{t-k} - \frac{k}{T} (x_T - x_0) \right]^2.$$ 

Thus the variance ratio can be estimated as

$$\hat{V}_k = \frac{\hat{\sigma}^2_k}{\hat{\sigma}^2_1}.$$ 

• $V_k$ can also be written as $V_k = 1 + 2 \sum_{j=1}^{k-1} \frac{k-j}{k} \rho_j$ so that the limiting variance ratio $V$ can be defined as

$$V = \lim_{k \to \infty} V_k = 1 + 2 \sum_{j=1}^{\infty} \rho_j.$$ 

• Since $\lim_{k \to \infty} \sigma^2_k = \frac{(\sum \psi_j)^2}{\sum \psi_j^2} \sigma^2_1 = (\sum \psi_j)^2 \sigma^2 = |A(1)|^2 \sigma^2$, $V$ can be written as $V = \frac{\sigma^2}{\sigma^2_1} |A(1)|^2$, which provides the link between the two persistence measure.

• By defining $R^2 = 1 - \frac{\sigma^2}{\sigma^2}$, the fraction of the variance that is predicable from knowledge of the past history of $\Delta x_t$, we have

$$A(1) = \sqrt{\frac{V}{1 - R^2}}.$$ 

the more predicable is $\Delta x_t$, the greater the difference between the two measures.
• The various unit root tests can have difficulties in detecting some important departure from a random walk, and the associated distribution of the test statistics tends to have awkward dependences on nuisance parameters.

• When the null hypothesis under examination is that the series is generated by a random walk with strict white noise normal increments, a test based on the variance ratio is preferred.

• Suppose \( x_t = \theta + x_{t-1} + a_t \). Lo and Mackinlay (1988, 1989) consider the test statistic

\[
M(k) = \frac{\sigma_k^2}{\sigma_1^2} - 1 = \hat{V}_k - 1
\]

and show that

\[
z_1(k) = M(k) \left[ \frac{2(2k - 1)(k - 1)}{3Tk} \right]^{-1/2} \sim N(0, 1).
\]

• If \( \{a_t\} \) satisfy Assumption I for non-parametric test, the following test statistic is robust to serial correlation and heteroscedasticity:

\[
z_2(k) = M(k) \cdot \Omega^{-1/2}(k), \text{ where }
\]

\[
\Omega(k) = \sum_{j=1}^{k-1} \left[ \frac{2(k - j)}{k} \right] \delta_j \text{ and } \delta_j = \frac{\Sigma_{j+1}^{T} \alpha_{0t} \alpha_{jt}}{(\Sigma_{t=1}^{T} \alpha_{0t})^2},
\]

\[
\alpha_{jt} = (x_{t-j} - x_{t-j-1} - \frac{1}{T}(x_T - x_0))^2.
\]
• Large-sample normal approximation works well when \( k \) is small and \( T \) is large. It can become unsatisfactory for large \( k \) because the empirical distribution of \( M(k) \) is highly skewed in these circumstances.

• Richardson and Stock (1989) consider a different perspective in which \( k \) is allowed to tend asymptotically to a non-zero fraction \( (\delta) \) of \( T \) i.e. \( k/T \to \delta \).

• Under this asymptotic theory, \( M(k) \) has a limiting distribution that is not normal

\[
M(k) \Rightarrow \frac{1}{\delta} \int_{\delta}^{1} Y(r)^2 dr,
\]

where \( Y(r) = W(r) - W(r - \delta) - \delta W(1) \).

• Richardson and Stock argue that the \( k/T \to \delta \) theory provides a much better approximation to the finite sample distribution of \( M(k) \) than does the fixed \( k \) theory.

• The limiting distribution is valid even under non-normality and certain form of heteroscedasticity.

• Lo and MacKinlay (1989) find that the power of the variance ratio test is comparable in power to \( \tau_\tau \) when \( x_t \) is trend stationary.
Example 3.6 Persistence and Mean Reversion in UK Stock Prices

- In example 2.6, we fitted an $ARIMA(3, 1, 0)$ to the logarithms of the *FTA All Share* Index with

$$\phi(B) = 1 - 0.152B + 0.140B^2 - 0.114B^3.$$  

- Thus $A(1) = 1/0.874 = 1.144$, which provide some evidence in favor of *mean aversion*, whereby a series will continue to diverge from its previously forecasted value following a shock.

- Cochrane (1988) criticize the use of fitted $ARIMA$ models for constructing the long-run measure $A(1)$ because they are designed to capture short-run dynamics.

- Furthermore, ARMA processes were unable to provide reliable estimates of $A(1)$ here because of the presence of approximate common factors in the AR and MA polynomials.

- Table 3.1 presents $M(k)$ statistics for a sequence of $k$ values associated with ‘long-difference’ of prices of between one and eight years.

- Also provided are the p-values using normal approximation and simulated upper tail percentiles using Richardson and Stock $k/T \rightarrow \delta$ asymptotic theory.
• Using either distribution, there is evidence to reject the random walk null at low levels of significance for values of $k > 48$ with smaller marginal significance level for the $k/T \rightarrow \delta$ distribution.

• We conclude that there is mean aversion in UK stock prices, with return being positively correlated at long horizons.

Table 3.1. **Variance ratio test statistics for UK stock prices (monthly data 1965-95)**

<table>
<thead>
<tr>
<th>$k$</th>
<th>$M$</th>
<th>$p_1$</th>
<th>95%</th>
<th>97%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.39</td>
<td>0.024</td>
<td>0.34</td>
<td>0.43</td>
<td>0.54</td>
</tr>
<tr>
<td>24</td>
<td>0.42</td>
<td>0.071</td>
<td>0.47</td>
<td>0.59</td>
<td>0.75</td>
</tr>
<tr>
<td>36</td>
<td>0.44</td>
<td>0.107</td>
<td>0.52</td>
<td>0.69</td>
<td>0.88</td>
</tr>
<tr>
<td>48</td>
<td>0.45</td>
<td>0.136</td>
<td>0.58</td>
<td>0.80</td>
<td>1.07</td>
</tr>
<tr>
<td>60</td>
<td>0.68</td>
<td>0.069</td>
<td>0.67</td>
<td>0.90</td>
<td>1.17</td>
</tr>
<tr>
<td>72</td>
<td>0.92</td>
<td>0.034</td>
<td>0.69</td>
<td>0.95</td>
<td>1.29</td>
</tr>
<tr>
<td>84</td>
<td>1.01</td>
<td>0.032</td>
<td>0.68</td>
<td>0.97</td>
<td>1.33</td>
</tr>
<tr>
<td>96</td>
<td>1.16</td>
<td>0.023</td>
<td>0.70</td>
<td>1.05</td>
<td>1.45</td>
</tr>
</tbody>
</table>

*Note:* $p_1(\cdot)$ denotes the probability under the null hypothesis of observing a larger variance ratio than that observed using the asymptotic $N(0, 1)$ distribution. 95%, 97.5%, 99% are percentiles of the empirical distribution of $M(k)$ computed under the $k/T \rightarrow \delta$ asymptotic theory using $NID(0, 1)$ returns with 5000 replications for each $k$. 
Fractional Integration and Long Memory

• In the analysis of financial time series, we usually consider the order of differencing, $d$, is either 0 or 1.
  
  $- x_t \sim I(1)$: The ACF declines linearly.
  
  $- x_t \sim I(0)$: The ACF declines exponentially so that observations separated by a long time span may be regarded as independent.

• Many empirically observed time series appeared to satisfy the assumption of stationarity (perhaps after some differencing transformation) but it exhibits dependence between distant observations.

• This is well known in hydrology, e.g. Hurst effect (Mandlebrot and Wallis, 1969).

• Many economic time series are characterized by the tendency for large values to be followed by large values of the same sign such that the series seem to go through a succession of cycles even including long cycles whose length is comparable to the total sample size.

• This leads to an extension of $ARMA$ class to model long-term persistence.
ARFIMA: Fractionally Integrated Processes

- Consider real $d > -1$,

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$$

$$= 1 - dB + \frac{d(d-1)}{2!} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \cdots$$

- Fractional white noise: $ARFIMA(0, d, 0)$ process

$$(1 - B)^d x_t = a_t.$$

is the discrete analog of fractional Brownian motion just like random walk is the discrete time analog of Brownian motion.

- How does the ARFIMA model incorporate long memory behavior?

- For non-integer values of $d$, ACF of $x_t$ declines hyperbolically to zero. The autocorrelations are given by $\rho_k = \Gamma k^{2d-1}$, $\Gamma$ is the ratio of two gamma functions. Thus the autocorrelation exhibits a hyperbolic decay and the speed of which depends on $d$.

- The process is weakly stationary for $d < 0.5$.

- For $d \geq 0.5$, $Var(x_t) = \infty$ so the process is non-stationary.

- But Robinson (1994) refers to it as being ‘less non-stationary’ than a unit root process, so smoothly bridging the gulf between $I(0)$ and $I(1)$ processes.
• The process is invertible for $d > -0.5$.
• Both the $\pi$- and $\phi$—weights exhibit slow hyperbolic decay for large $k$.
• The same properties are displayed by the more general $ARIMA(p, d, q)$ process
  \[ \phi(B)(1 - B)^d x_t = \theta(B)a_t. \]
• The impulse response function is defined by
  \[ \Delta x_t = (1 - B)^d \phi(B)^{-1}\theta(B)a_t = A(B)a_t. \]
• From Baillie (1996), $(1 - B)^{1-d}=0$ for $d < 1$, so that any $ARIMA$ process is trend reverting since $A(1) = 0$.
• The intuition behind the concept of long memory and the fractional difference emerge more clearly in the frequency domain.

  Spectral Density

• Consider a stochastic process $X_t$ with the following representation
  \[ X_t = \sum_{j=1}^{n} a_j \exp(it\omega_j), \]
  where
  $\omega_j, -\pi < \omega_1 < \cdots < \omega_n < \pi$ are Fourier frequencies
  $a_j$ are uncorrelated random Fourier coefficients.
The preceding process is stationary with
\[ EX_t = 0, \quad \gamma_k = EX_t \bar{X}_{t-k} = \sum_{j=1}^{n} Ea_j^2 \exp(ik\omega_j). \]

Hence \[ \gamma_k = \int_{(-\pi,\pi]} \exp(ik\omega) dF(\omega), \quad (7) \]
where \( F(\omega) \) is the spectral distribution with spectral density \( f(\omega_j) = E|a_j|^2 \).

The remarkable feature of the preceding example is that every zero-mean stationary process permits the spectral representation
\[ X_t = \int_{(-\pi,\pi]} \exp(it\omega) dZ(\omega), \]
where \( Z \) is an orthogonal increment process.

Furthermore the autocovariance admits a similar spectral representation as (7).

When \( f(\omega) \) yields high value, the process exhibits cyclical behavior at frequency \( \omega \) with the period of one cycle equaling \( 2\pi/\omega \) time units.

The series \( x_t \) will display long memory if its spectral density, \( f(\omega) \), increase without limit as the frequency \( \omega \) tends to zero.

If \( x_t \) is ARFIMA, then \( f_x(\omega) \) behaves like \( \omega^{-2d} \) as \( \omega \rightarrow 0 \). So \( d \) parameterizes the low-frequency behavior.
Test for Fractional Difference

Classical approach to detect the presence of long-term memory can be found in Hurst (1951), Mandelbrot (1972), which employ the $R/S$ statistic

$$R_0 = \hat{\sigma}_0^{-1} \left[ \max_{1 \leq i \leq T} \sum_{t=1}^{i} (x_t - \bar{x}) - \min_{1 \leq i \leq T} \sum_{t=1}^{i} (x_t - \bar{x}) \right], \quad (8)$$

where $\hat{\sigma}_0^2 = T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2$.

- The first term in the brackets is the maximum (over $T$) of the partial sums of the first $i$ deviation of $x_i$ from the sample mean. Since the sum of all $T$ deviations from mean is zero, the maximum is always non-negative. The second term is the minimum of the same sequence, and hence is always non-positive. Therefore $R_0 \geq 0$.

- Although it has long been established that the R/S statistic has the ability to detect long-range dependence, it is sensitive to short-range dependence.

- Any incompatibility between the data and the predicted behavior of the R/S statistic under the null hypothesis of no long-run dependence need not come from long-term memory, but may be merely a symptom of short-term autocorrelation.
Thus Lo (1991) modified the R/S statistic in which short-run dependence is incorporated into its denominator, which becomes the square root of a consistent estimator of the variance of the partial sum in

$$R_q = \hat{\sigma}_q^{-1} \left[ \max_{1 \leq i \leq T} \sum_{t=1}^{i} (x_t - \bar{x}) - \min_{1 \leq i \leq T} \sum_{t=1}^{i} (x_t - \bar{x}) \right], \quad (9)$$

where for $q < T$

$$\hat{\sigma}_q^2 = \hat{\sigma}_0^2 \left( 1 + \frac{2}{T} \sum_{j=1}^{q} w_{qj} r_j \right), \quad w_{qj} = 1 - \frac{j}{q + 1},$$

and $r_j$ is the sample autocorrelations.

- The asymptotic distribution of $T^{-1/2} R_q$ is the same as that of the range of Brownian bridge in unit interval (see Lo, 1991).

- This test is consistent against a class of long-range dependent alternatives that include all ARFIMA($p, d, q$) models with $-0.5 \leq d \leq 0.5$.

- However, the choice of $q$ is an unresolved issue.

- There is evidence that if the distribution of $x_t$ is fat-tailed, then the sampling distribution of $R_q$ is shifted to the left relative to the asymptotic distribution.

- Lo thus argue that the R/S approach may be best regarded as a kind of portmanteau test that may complement, and come prior to a more comprehensive analysis of long-range dependence.
LM Tests of $d = 0$:

- An obvious approach to testing for fractional differencing is to construct tests against the null of either $d = 1$ or $d = 0$.

- The ADF test and non-parametric test of $d = 1$ are consistent against fractional $d$ alternative, although the power of the tests grows more slowly as $d$ diverges from unity than with the divergence of the AR parameter from unity.

- Lee and Schmidt (1996) show that the $\eta$ statistic of Kwiatkowski et al (1992) for testing the null $d = 0$ are consistent against fractional $d$ alternative in the range $-0.5 < d < 0.5$, and their power compares favorably to Lo’s modified R/S statistic.

- Alternatively, we can construct tests based on the residuals from fitting an $ARIMA(p, 0, q)$ model to $x_t$.

- Suppose that the fitted model is

$$\hat{\phi}(B)x_t = \hat{\theta}(B)\hat{a}_t$$

- Agiakloglou and Newbold (1994) derive LM test of $d = 0$ as the $t$-ratio on $\delta$ in the following regression.

$$\hat{a}_t = \sum_{i=1}^{p} \beta_i W_{t-i} + \sum_{j=1}^{q} \gamma_j W_{t-j} + \delta K_t(m) + u_t$$
where $\hat{\theta}(B)W_t = x_t$, $\hat{\theta}(B)Z_t = \hat{a}_t$, and $K_t(m) = \sum_{j=1}^{m} j^{-1}\hat{a}_{t-j}$.

- For the preceding LM test, low power is a particular problem when $p$ and $q$ are positive.

- Mean estimation for long memory process is a general problem as the sample mean is a poor estimate of the true mean.

- Indeed, Agiakloglou and Newbold (1993) find that the SACF of fractional white noise (when $d > 0$) is a severely biased estimator of the true ACF, so that it is very difficult to detect long memory behavior from the SACF’s of moderate length series.

**Estimation of $d$: GPH method**


- The spectral density of $x_t$ is given by

$$f_x(w) = |1 - \exp(-iw)|^{-2d}f_w(w) = [4\sin^2(w/2)]^{-d}f_w(w),$$

where $f_w(w)$ is the spectral density of $w_t = (1-B)^d x_t$.

- It then follows that

$$\log[f_x(w)] = \log[f_W(w)] - d\log[4\sin^2(w/2)].$$
• This leads GPH to estimate $d$ as (minus) the slope estimator of the regression of the periodogram $I(w_j)$ on a constant and $\log[4\sin^2(w/2)]$ at frequencies $w_j = 2\pi j/T$, $j = 1, \cdots, K = [T^{1/2}]$.

• The periodogram is defined to be

$$I(\omega_j) = T^{-1}\left|\sum_{t=1}^{T} x_t \exp(-it\omega_j)\right|^2.$$ 

In a sense, periodogram is the sample version of spectral density. Note that

$$\sum_{t=1}^{T} x_j^2 = \sum_{j} I(\omega_j).$$

• Although PGH estimator is consistent, asymptotically normal, and robust to non-normality, it is severely affected by autocorrelation in $w_t$. That is, $x_t$ is an ARFIMA process rather than fractional white noise.

• It is both biased and inefficient in these circumstances. In particular, when there are large positive AR or MA roots the estimator of $d$ is seriously biased upwards, so that the null of $d = 0$ would be rejected far too often.

• Attention has therefore focused on joint MLE of all parameters in the $ARFIMA(p, d, q)$ models.
Example 3.7 Exchange Rate and Stock Returns

- In Example 3.1, it is confirmed that Dollar/Sterling exchange rate contains a unit root, whereas in Example 3.2, we confirm that is also the case for FTA All Share Index.

- We now consider whether the differences of the two series, the returns, are really stationary or they exhibit long memory.

  Dollar/Sterling Exchange

- Applying the modified R/S statistic with \( q = \lfloor T^{1/4} \rfloor = 9 \) (recommendation of Lo, 1991) to the exchange rate difference, we obtain

  \[
  T^{-1/2} R_9 = 1.692, \text{ with } 95\% \text{ CI (0.809, 1.862).}
  \]

  One cannot reject the hypothesis that exchange rate returns are short memory.

- Using the residuals from an ARIMA(1, 1, 0) model, the LM test yields \( t \)-ratios for \( \delta \) equaling 1.03, 1.23, 1.30 and 1.21 for \( m \) set equal to 25, 50, 75 and 100 respectively.

- The GPH estimate is such that \( \hat{d} = -0.07 \pm 0.08 \) with \( K = \lfloor T^{1/2} \rfloor = 22 \).

- All results confirm that the exchange rate difference is not long memory.
FTA All Share Index

• The modified R/S with $q = 4$ yields $T^{-1/2}R_4 = 2.090$, which is significant.

• This finding of long memory is echoed by the GPH estimate of $\hat{d} = 0.39 \pm 0.19$ using $K = 19$.

• However, LM $t$-ratios were never remotely significant for a wide range of $m$ values. This is consistent with the simulation results that the power of this test is very week when the sample mean (the drift of the index) has to be estimated.

• The empirical evidence is thus consistent with the *FTA All Share Index* being an $I(1.4)$ instead of $I(1)$.

Daily returns for the S&P 500 Index

• We investigate the daily returns from January 1928 to August 1991, a total of $T = 17,054$ observations.

• The GPH estimate was $\hat{d} = 0.11 \pm 0.06$, so that there is little evidence that the series is long memory.

• However, for the squared returns series we obtain $\hat{d} = 0.56$, while for absolute returns we have $\hat{d} = 0.73$.

• Thus simple nonlinear transformation of returns do appear to be long memory, and this is also found to be the case for wide variety of other financial time series.