Wavelet Methods for Time Series Analysis

Part VIII: Wavelet-Based Analysis and Synthesis of Long Memory Processes

- DWT well-suited for long memory processes (LMPs)
- basic idea: DWT approximately decorrelates LMPs
- on synthesis side, leads to DWT-based simulation of LMPs
- on analysis side, leads to wavelet-based maximum likelihood and least squares estimators for LMP parameters, along with a test for homogeneity of variance
Wavelets and Long Memory Processes: I

• wavelet filters are approximate band-pass filters, with nominal pass-bands \([1/2^j+1, 1/2^j]\) (called \(j\)th ‘octave band’)
• suppose \(\{X_t\}\) has \(S_X(\cdot)\) as its spectral density function (SDF)
• statistical properties of \(\{W_{j,t}\}\) are simple if \(S_X(\cdot)\) has simple structure within \(j\)th octave band
• example: fractionally differenced (FD) process

\[(1 - B)^\delta X_t = \varepsilon_t,\]

(where \(B\) is the backward shift operator such that \((1 - B)X_t = X_t - X_{t-1}\)) having SDF

\[S_X(f) = \sigma_\varepsilon^2/[4\sin^2(\pi f)]^\delta\]
Wavelets and Long Memory Processes: II

- FD process controlled by two parameters: $\delta$ and $\sigma^2\varepsilon$
- for small $f$, have $S_X(f) \approx C|f|^{-2\delta}$; i.e., a power law
- log($S_X(f)$) vs. log($f$) is approximately linear with slope $-2\delta$
- for large $\tau_j$, the wavelet variance at scale $\tau_j$, namely $\nu^2_X(\tau_j)$, satisfies $\nu^2_X(\tau_j) \approx C' \tau_j^{2\delta-1}$
- log ($\nu^2_X(\tau_j)$) vs. log ($\tau_j$) is approximately linear, slope $2\delta - 1$
- approximately ‘self-similar’ (or ‘fractal’)
- stationary ‘long memory’ process (LMP) if $0 < \delta < 1/2$: correlation between $X_t$ and $X_{t+\tau}$ dies down slowly as $\tau$ increases
Wavelets and Long Memory Processes: III

• power law model ubiquitous in physical sciences
  – voltage fluctuations across cell membranes
  – traffic fluctuations on an expressway
  – impedance fluctuations in geophysical borehole
  – fluctuations in the rotation of the earth
  – X-ray time variability of galaxies

• DWT well-suited to study FD process and other LMPs
  – ‘self-similar’ filters used on ‘self-similar’ processes
  – key idea: DWT approximately decorrelates LMPs
DWT of a Long Memory Process: I

- realization of an FD(0.4) time series $X$ along with its sample autocorrelation sequence (ACS): for $\tau \geq 0$,

$$\hat{\rho}_{X,\tau} \equiv \frac{\sum_{t=0}^{N-1-\tau} X_t X_{t+\tau}}{\sum_{t=0}^{N-1} X_t^2}$$

- note that ACS dies down slowly
• LA(8) DWT of FD(0.4) series and sample ACSs for each $W_j$ & $V_7$, along with 95% confidence intervals for white noise
MODWT of a Long Memory Process

\[ \tilde{V}_7 \] \[ \tilde{W}_7 \] \[ \tilde{W}_6 \] \[ \tilde{W}_5 \] \[ \tilde{W}_4 \] \[ \tilde{W}_3 \] \[ \tilde{W}_2 \] \[ \tilde{W}_1 \]

- LA(8) MODWT of FD(0.4) series & sample ACSs for MODWT coefficients, none of which are approximately uncorrelated
DWT of a Long Memory Process: III

• in contrast to $X$, ACSs for $W_j$ consistent with white noise

• variance of $W_j$ increases with $j$ – to see why, note that

$$\text{var} \{W_{j,t}\} = \int_{-1/2}^{1/2} \mathcal{H}_j(f) S_X(f) \, df$$

$$\approx 2 \int_{1/2j+1}^{1/2j} 2^j S_X(f) \, df$$

$$= \frac{1}{2^j - 2^{j+1}} \int_{1/2j+1}^{1/2j} S_X(f) \, df \equiv C_j,$$

where $C_j$ is average value of $S_X(\cdot)$ over $[1/2^{j+1}, 1/2^j]$

• for FD process, can argue that $C_j \approx S_X(1/2^{j+1/2})$, where $1/2^{j+1/2}$ is midpoint of interval $[1/2^{j+1}, 1/2^j]$
plot shows \( \hat{\text{var}} \{ W_{j,t} \} \) (circles) & \( S_X(1/2^{j+1/2}) \) (curve) versus \( 1/2^{j+1/2} \), along with 95\% confidence intervals for \( \text{var} \{ W_{j,t} \} \)

observed \( \hat{\text{var}} \{ W_{j,t} \} \) agrees well with theoretical \( \text{var} \{ W_{j,t} \} \)
Correlations Within a Scale and Between Two Scales

- let \( \{s_{X,\tau}\} \) denote autocovariance sequence (ACVS) for \( \{X_t\} \); i.e., \( s_{X,\tau} = \text{cov} \{X_t, X_{t+\tau}\} \)
- let \( \{h_{j,l}\} \) denote equivalent wavelet filter for \( j \)th level
- to quantify decorrelation, can write

\[
\text{cov} \{W_{j,t}, W_{j',t'}\} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l}h_{j',l'}s_{X,2^j(t+1)-l-2^{j'}(t'+1)+l',2^{j'}(t'+1)+l'}
\]

from which we can get ACVS (and hence within-scale correlations) for \( \{W_{j,t}\} \):

\[
\text{cov} \{W_{j,t}, W_{j,t+\tau}\} = \sum_{m=-(L_j-1)}^{L_j-1} s_{X,2^j\tau+m} \sum_{l=0}^{L_j-|m|-1} h_{j,l}h_{j,l+|m|}
\]
Correlations Within a Scale

- correlations between $W_{j,t}$ and $W_{j,t+\tau}$ for an FD(0.4) process
- correlations within scale are slightly smaller for Haar
- maximum magnitude of correlation is less than 0.2
Correlations Between Two Scales: I

\begin{align*}
  j' &= 2 \quad j' &= 3 \quad j' &= 4 \\
  j &= 1 \quad j &= 2 \quad j &= 3 \\
  \tau &
\end{align*}

- correlation between Haar wavelet coefficients $W_{j,t}$ and $W_{j',t'}$
  from FD(0.4) process and for levels satisfying $1 \leq j < j' \leq 4$
**Correlations Between Two Scales: II**

<table>
<thead>
<tr>
<th>$j' = 2$</th>
<th>$j' = 3$</th>
<th>$j' = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$j = 1$</td>
<td>$j = 2$</td>
<td>$j = 3$</td>
</tr>
</tbody>
</table>

- same as before, but now for LA(8) wavelet coefficients
- correlations between scales decrease as $L$ increases
Wavelet Domain Description of FD Process

- DWT acts as a decorrelating transform for FD process (also true for fractional Gaussian noise, pure power law etc.)
- wavelet domain description is simple
- wavelet coefficients within a given scale are approximately uncorrelated (refinement: assume 1st order autoregressive model)
- wavelet coefficients have a scale-dependent variance, but these variances are controlled by the two FD parameters ($\delta$ and $\sigma_\epsilon^2$)
- wavelet coefficients between scales are also approximately uncorrelated (approximation improves as filter width $L$ increases)
DWT-Based Simulation

• properties of DWT of FD processes lead to schemes for simulating time series \( \mathbf{X} \equiv [X_0, \ldots, X_{N-1}]^T \) with zero mean and with a multivariate Gaussian distribution

• with \( N = 2^J \), recall that \( \mathbf{X} = \mathbf{W}^T \mathbf{W} \), where

\[
\mathbf{W} = \begin{bmatrix}
\mathbf{W}_1 \\
\mathbf{W}_2 \\
\vdots \\
\mathbf{W}_j \\
\vdots \\
\mathbf{W}_J \\
\mathbf{V}_J
\end{bmatrix}
\]
Basic DWT-Based Simulation Scheme

• assume $W$ to contain $N$ uncorrelated Gaussian (normal) random variables (RVs) with zero mean

• assume $W_j$ to have variance $C_j \approx S_X(1/2^j+\frac{1}{2})$

• assume single RV in $V_J$ to have variance $C_{J+1}$ (see textbook for details about how to set $C_{J+1}$)

• approximate FD time series $X$ via $Y \equiv W^T \Lambda^{1/2}Z$, where
  - $\Lambda^{1/2}$ is $N \times N$ diagonal matrix with diagonal elements
    \[
    C_1^{1/2}, \ldots, C_1^{1/2}, C_2^{1/2}, \ldots, C_2^{1/2}, \ldots, C_{J-1}^{1/2}, C_{J-1}^{1/2}, C_J^{1/2}, C_J^{1/2}, C_{J+1}^{1/2}
    \]
    \[
    \begin{array}{ll}
    \frac{N}{2} \text{ of these} & \frac{N}{4} \text{ of these} \\
    \end{array}
    \]
    \[
    \begin{array}{ll}
    2 \text{ of these} & \end{array}
    \]
  - $Z$ is vector of deviations drawn from a Gaussian distribution with zero mean and unit variance
Refinements to Basic Scheme: I

- covariance matrix for approximation $\mathbf{Y}$ does not correspond to that of a stationary process
- recall $\mathcal{W}$ treats $\mathbf{X}$ as if it were circular
- let $\mathbf{T}$ be $N \times N$ ‘circular shift’ matrix:

\[
\mathbf{T} \begin{bmatrix}
  Y_0 \\
  Y_1 \\
  Y_2 \\
  Y_3 \\
\end{bmatrix} = \begin{bmatrix}
  Y_1 \\
  Y_2 \\
  Y_3 \\
  Y_0 \\
\end{bmatrix}; \quad \mathbf{T}^2 \begin{bmatrix}
  Y_0 \\
  Y_1 \\
  Y_2 \\
  Y_3 \\
\end{bmatrix} = \begin{bmatrix}
  Y_2 \\
  Y_3 \\
  Y_0 \\
  Y_1 \\
\end{bmatrix}; \quad \text{etc.}
\]

- let $\kappa$ be uniformly distributed over $0, \ldots, N - 1$
- define $\tilde{\mathbf{Y}} \equiv \mathbf{T}^\kappa \mathbf{Y}$
- $\tilde{\mathbf{Y}}$ is stationary with ACVS given by, say, $s_{\tilde{Y}, \tau}$
Refinements to Basic Scheme: II

- Q: how well does \( \{s_{\tilde{Y},\tau}\} \) match \( \{s_{X,\tau}\} \)?

- due to circularity, find that \( s_{\tilde{Y},N-\tau} = s_{\tilde{Y},\tau} \) for \( \tau = 1, \ldots, N/2 \)

- implies \( s_{\tilde{Y},\tau} \) cannot approximate \( s_{X,\tau} \) well for \( \tau \) close to \( N \)

- can patch up by simulating \( \tilde{Y} \) with \( M > N \) elements and then extracting first \( N \) deviates (\( M = 4N \) works well)
plot shows true ACVS $\{s_{X,\tau}\}$ (thick curves) for FD(0.4) process and wavelet-based approximate ACVSs $\{s_{\tilde{Y},\tau}\}$ (thin curves) based on an LA(8) DWT in which an $N = 64$ series is extracted from $M = N$, $M = 2N$ and $M = 4N$ series
Example and Some Notes

- simulated FD(0.4) series (LA(8), \(N = 1024\) and \(M = 4N\))
- notes:
  - can form realizations faster than best exact method
  - efficient ‘real-time’ simulation of extremely long time series (e.g, \(N = 2^{30} = 1,073,741,824\) or even longer)
  - effect of random circular shifting is to render time series non-Gaussian (a Gaussian mixture model)
MLEs of FD Parameters: I

• FD process depends on 2 parameters, namely, $\delta$ and $\sigma^2_{\varepsilon}$:

$$S_X(f) = \frac{\sigma^2_{\varepsilon}}{[4 \sin^2(\pi f)]^\delta}$$

• given $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ with $N = 2^J$, suppose we want to estimate $\delta$ and $\sigma^2_{\varepsilon}$

• if $\mathbf{X}$ is stationary (i.e. $\delta < 1/2$) and multivariate Gaussian, can use the maximum likelihood (ML) method
MLEs of FD Parameters: II

- definition of Gaussian likelihood function:

\[
L(\delta, \sigma^2_\epsilon \mid X) \equiv \frac{1}{(2\pi)^{N/2}|\Sigma_X|^{1/2}} e^{-X^T \Sigma_X^{-1} X / 2}
\]

where \(\Sigma_X\) is covariance matrix for \(X\), with \((s, t)\)th element given by \(s_X, s_t\), and \(|\Sigma_X| & \Sigma_X^{-1}\) denote determinant & inverse

- ML estimators of \(\delta\) and \(\sigma^2_\epsilon\) maximize \(L(\delta, \sigma^2_\epsilon \mid X)\) or, equivalently, minimize

\[
-2 \log (L(\delta, \sigma^2_\epsilon \mid X)) = N \log (2\pi) + \log (|\Sigma_X|) + X^T \Sigma_X^{-1} X
\]

- exact MLEs computationally intensive, mainly because of the need to invert \(\Sigma_X\) and compute its determinant

- good approximate MLEs of considerable interest
MLEs of FD Parameters: III

• key ideas behind first wavelet-based approximate MLEs
  – have seen that we can approximate FD time series $X$ by $Y = \mathcal{W}^T \Lambda^{1/2} Z$, where $\Lambda^{1/2}$ is a diagonal matrix, all of whose diagonal elements are positive
  – since covariance matrix for $Z$ is $I_N$, Equation (262c) says covariance matrix for $Y$ is
    \[ \mathcal{W}^T \Lambda^{1/2} I_N \left( \mathcal{W}^T \Lambda^{1/2} \right)^T = \mathcal{W}^T \Lambda^{1/2} \Lambda^{1/2} \mathcal{W} = \mathcal{W}^T \Lambda \mathcal{W} \equiv \tilde{\Sigma}_X, \]
    where $\Lambda \equiv \Lambda^{1/2} \Lambda^{1/2}$ is also diagonal
  – can consider $\tilde{\Sigma}_X$ to be an approximation to $\Sigma_X$
• leads to approximation of log likelihood:
  \[ -2 \log \left( L(\delta, \sigma_\varepsilon^2 | X) \right) \approx N \log (2\pi) + \log (|\tilde{\Sigma}_X|) + X^T \tilde{\Sigma}_X^{-1} X \]
MLEs of FD Parameters: IV

• Q: so how does this help us?
  - easy to invert $\tilde{\Sigma}_X$:

$$\tilde{\Sigma}_X^{-1} = \left( \mathcal{W}^T \Lambda \mathcal{W} \right)^{-1} = \left( \mathcal{W} \right)^{-1} \Lambda^{-1} \left( \mathcal{W}^T \right)^{-1} = \mathcal{W}^T \Lambda^{-1} \mathcal{W},$$

where $\Lambda^{-1}$ is another diagonal matrix, leading to

$$X^T \tilde{\Sigma}^{-1}_X X = X^T \mathcal{W}^T \Lambda^{-1} \mathcal{W} X = \mathcal{W}^T \Lambda^{-1} \mathcal{W}$$

  - easy to compute the determinant of $\tilde{\Sigma}_X$:

$$|\tilde{\Sigma}_X| = |\mathcal{W}^T \Lambda \mathcal{W}| = |\Lambda \mathcal{W} \mathcal{W}^T| = |\Lambda I_N| = |\Lambda| \cdot |I_N| = |\Lambda|,$$

and the determinant of a diagonal matrix is just the product of its diagonal elements
MLEs of FD Parameters: V

• define the following three functions of $\delta$:

$$C'_{j}(\delta) \equiv \int_{1/2j+1}^{1/2j} \frac{2^{j+1}}{\sin^2(\pi f)^\delta} df \approx \int_{1/2j+1}^{1/2j} \frac{2^{j+1}}{(2\pi f)^{2\delta}} df$$

$$C'_{J+1}(\delta) \equiv \frac{N\Gamma(1 - 2\delta)}{\Gamma^2(1 - \delta)} - \sum_{j=1}^{J} N\frac{N}{2^{j}} C'_{j}(\delta)$$

$$\sigma^2_{\varepsilon}(\delta) \equiv \frac{1}{N} \left( \frac{V_{J,0}^2}{C'_{J+1}(\delta)} + \sum_{j=1}^{J} \frac{1}{C'_{j}(\delta)} \sum_{t=0}^{N/2^{j}-1} W_{j,t}^2 \right)$$
MLEs of FD Parameters: VI

- wavelet-based approximate MLE $\tilde{\delta}$ for $\delta$ is the value that minimizes the following function of $\delta$:

$$\tilde{l}(\delta \mid X) \equiv N \log(\sigma_{\tilde{\varepsilon}}^2(\delta)) + \log(C'_{J+1}(\delta)) + \sum_{j=1}^{J} \frac{N}{2^j} \log(C'_j(\delta)),$$

- once $\tilde{\delta}$ has been determined, MLE for $\sigma_{\tilde{\varepsilon}}^2$ is given by $\sigma_{\tilde{\varepsilon}}^2(\tilde{\delta})$
- computer experiments indicate scheme works quite well
LSEs of FD Parameters

- one alternative to MLEs are least square estimators (LSEs)
  - recall that, for large $\tau$ and for $\beta = 2\delta - 1$,
    \[
    \log (\nu_X^2(\tau_j)) \approx \zeta + \beta \log (\tau_j)
    \]
  - suggests determining $\delta$ by regressing $\log (\hat{\nu}_X^2(\tau_j))$ on $\log (\tau_j)$ over range of $\tau_j$
  - weighted LSE takes into account fact that variance of $\log (\hat{\nu}_X^2(\tau_j))$ depends upon scale $\tau_j$ (increases as $\tau_j$ increases)
Homogeneity of Variance: I

• because DWT decorrelates LMPs, nonboundary coefficients in $W_j$ should resemble white noise; i.e., $\text{cov}\{W_{j,t}, W_{j,t'}\} \approx 0$ when $t \neq t'$, and $\text{var}\{W_{j,t}\}$ should not depend upon $t$

• can test for homogeneity of variance in $X$ using $W_j$ at each level $j$

• suppose $U_0, \ldots, U_{N-1}$ are independent normal RVs with $E\{U_t\} = 0$ and $\text{var}\{U_t\} = \sigma_t^2$

• want to test null hypothesis

$$H_0 : \sigma_0^2 = \sigma_1^2 = \cdots = \sigma_{N-1}^2$$

• can test $H_0$ versus a variety of alternatives, e.g.,

$$H_1 : \sigma_0^2 = \cdots = \sigma_k^2 \neq \sigma_{k+1}^2 = \cdots = \sigma_{N-1}^2$$

using normalized cumulative sum of squares
Homogeneity of Variance: II

- to define test statistic $D$, start with

$$P_k \equiv \frac{\sum_{j=0}^{k} U_j^2}{\sum_{j=0}^{N-1} U_j^2}, \quad k = 0, \ldots, N - 2$$

and then compute $D \equiv \max (D^+, D^-)$, where

$$D^+ \equiv \max_{0 \leq k \leq N-2} \left( \frac{k + 1}{N - 1} - P_k \right) \quad \& \quad D^- \equiv \max_{0 \leq k \leq N-2} \left( P_k - \frac{k}{N - 1} \right)$$

- can reject $H_0$ if observed $D$ is ‘too large,’ where ‘too large’ is quantified by considering distribution of $D$ under $H_0$

- need to find critical value $x_\alpha$ such that $P[D \geq x_\alpha] = \alpha$ for, e.g., $\alpha = 0.01, 0.05$ or 0.1
Homogeneity of Variance: III

- once determined, can perform $\alpha$ level test of $H_0$:
  - compute $D$ statistic from data $U_0, \ldots, U_{N-1}$
  - reject $H_0$ at level $\alpha$ if $D \geq x_\alpha$
  - fail to reject $H_0$ at level $\alpha$ if $D < x_\alpha$

- can determine critical values $x_\alpha$ in two ways
  - Monte Carlo simulations
  - large sample approximation to distribution of $D$:
    \[ P\left(\left(\frac{N}{2}\right)^{1/2}D \geq x\right) \approx 1 + 2 \sum_{l=1}^{\infty}(-1)^l e^{-2l^2x^2} \]
    (reasonable approximation for $N \geq 128$)
Homogeneity of Variance: IV

• idea: given time series \( \{X_t\} \), compute \( D \) using nonboundary wavelet coefficients \( W_{j,t} \) (there are \( M'_j \equiv N_j - L'_j \) of these):

\[
\mathcal{P}_k \equiv \frac{\sum_{t=L'_j}^{L'_j+\tilde{k}} W_{j,t}^2}{\sum_{t=L'_j}^{N_j-1} W_{j,t}^2}, \quad k = L'_j, \ldots, N_j - 2
\]

• if null hypothesis rejected at level \( j \), can use nonboundary MODWT coefficients to accurately locate change point based on

\[
\tilde{\mathcal{P}}_k \equiv \frac{\sum_{t=L_j-1}^{L_j-1+\tilde{k}} \tilde{W}_{j,t}^2}{\sum_{t=L_j-1}^{N-1} \tilde{W}_{j,t}^2}, \quad k = L_j - 1, \ldots, N - 2
\]

along with analogs \( \tilde{D}_k^+ \) and \( \tilde{D}_k^- \) of \( D_k^+ \) and \( D_k^- \)
Annual Minima of Nile River

- left-hand plot: annual minima of Nile River
- new measuring device introduced in year 715
- right: Haar $\hat{\nu}_X^2(\tau_j)$ before ($\times$’s) and after ($\circ$’s) year 715.5, with 95% confidence intervals based upon $\chi^2_{\eta_3}$ approximation
Example – Annual Minima of Nile River: II

- based upon last 512 values (years 773 to 1284), plot shows \( \tilde{l}(\delta \mid X) \) versus \( \delta \) for the first wavelet-based approximate MLE using the LA(8) wavelet (upper curve) and corresponding curve for exact MLE (lower)
  - wavelet-based approximate MLE is value minimizing upper curve: \( \tilde{\delta} \doteq 0.4532 \)
  - exact MLE is value minimizing lower curve: \( \hat{\delta} \doteq 0.4452 \)
Example – Annual Minima of Nile River: III

- using last 512 values again, variance of wavelet coefficients computed via LA(8) MLEs $\tilde{\delta}$ and $\sigma_\varepsilon^2(\tilde{\delta})$ (solid curve) as compared to sample variances of LA(8) wavelet coefficients (circles)
- agreement is almost too good to be true!
Example – Annual Minima of Nile River: IV

- Results of testing all Nile River minima for homogeneity of variance using the Haar wavelet filter with critical values determined by computer simulations

<table>
<thead>
<tr>
<th>$\tau_j$</th>
<th>$M'_j$</th>
<th>$D$</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year</td>
<td>331</td>
<td>0.1559</td>
<td>0.0945</td>
<td>0.1051</td>
<td>0.1262</td>
</tr>
<tr>
<td>2 years</td>
<td>165</td>
<td>0.1754</td>
<td>0.1320</td>
<td>0.1469</td>
<td>0.1765</td>
</tr>
<tr>
<td>4 years</td>
<td>82</td>
<td>0.1000</td>
<td>0.1855</td>
<td>0.2068</td>
<td>0.2474</td>
</tr>
<tr>
<td>8 years</td>
<td>41</td>
<td>0.2313</td>
<td>0.2572</td>
<td>0.2864</td>
<td>0.3436</td>
</tr>
</tbody>
</table>

- Can reject null hypothesis of homogeneity of variance at level of significance 0.05 for scales $\tau_1$ & $\tau_2$, but not at larger scales.
Example – Annual Minima of Nile River: V

- Nile River minima (top plot) along with curves (constructed per Equation (382)) for scales $\tau_1$ & $\tau_2$ (middle & bottom) to identify change point via time of maximum deviation (vertical lines denote year 715)
Summary

• wavelets approximately decorrelate LMPs
• leads to practical and flexible schemes for simulating LMPs
• also leads to schemes for estimating parameters of LMPs
  — approximate maximum likelihood estimators
  — weighted least squares estimator
• can also devise wavelet-based tests for
  — homogeneity of variance
  — trends (see Section 9.4 & Craigmile et al., *Environmetrics*, 15, 313–35, 2004, for details)