Wavelet Variance – Outline

• examples of time series to motivate discussion
• decomposition of sample variance using wavelets
• theoretical wavelet variance for stochastic processes
  – stationary processes
  – nonstationary processes with stationary differences
• sampling theory for Gaussian processes
• four examples, including use on time series with time-varying statistical properties
• summary
Examples: Time Series $X_t$ Versus Time Index $t$

(a) atomic clock frequency deviates (daily observations, $N = 1025$)
(b) subtidal sea level fluctuations (twice daily, $N = 8746$)
(c) Nile River minima (annual, $N = 663$)
(d) vertical shear in the ocean (0.1 meters, $N = 4096$)

- four series are visually different
- goal of time series analysis is to quantify these differences
Decomposing Sample Variance of Time Series

• one approach: quantify differences by analysis of variance
• let $X_0, X_1, \ldots, X_{N-1}$ represent time series with $N$ values
• let $\overline{X}$ denote sample mean of $X_t$’s: $\overline{X} \equiv \frac{1}{N} \sum_{t=0}^{N-1} X_t$
• let $\hat{\sigma}^2_X$ denote sample variance of $X_t$’s:

$$\hat{\sigma}^2_X \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2$$

• idea is to decompose (analyze, break up) $\hat{\sigma}^2_X$ into pieces that quantify how time series are different
• wavelet variance does analysis based upon differences between (possibly weighted) adjacent averages over scales
Empirical Wavelet Variance

• define empirical wavelet variance for scale \( \tau_j \equiv 2^{j-1} \) as

\[
\tilde{\nu}^2_X(\tau_j) \equiv \frac{1}{N} \sum_{t=0}^{N-1} \tilde{W}_{j,t}^2, \quad \text{where} \quad \tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N}
\]

• if \( N = 2^J \), obtain analysis (decomposition) of sample variance:

\[
\hat{\sigma}^2_X = \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \sum_{j=1}^{J} \tilde{\nu}^2_X(\tau_j)
\]

(if \( N \) not a power of 2, can analyze variance to any level \( J_0 \), but need additional component involving scaling coefficients)

• interpretation: \( \tilde{\nu}^2_X(\tau_j) \) is portion of \( \hat{\sigma}^2_X \) due to changes in averages over scale \( \tau_j \); i.e., ‘scale by scale’ analysis of variance
Example of Empirical Wavelet Variance

- wavelet variances for time series $X_t$ and $Y_t$ of length $N = 16$, each with zero sample mean and same sample variance

![Graphs of $X_t$ and $Y_t$ with wavelet variances $\hat{\nu}_X^2(\tau_j)$ and $\hat{\nu}_Y^2(\tau_j)$](image)
Second Example of Empirical Wavelet Variance

- top: part of subtidal sea level data (blue line shows scale of 16)

- bottom: empirical wavelet variances $\tilde{\nu}^2_X(\tau_j)$
- note: each $\tilde{W}_{j,t}$ associated with a portion of $X_t$, so $\tilde{W}_{j,t}^2$ versus $t$ offers time-based decomposition of $\tilde{\nu}^2_X(\tau_j)$
Theoretical Wavelet Variance: I

• now assume $X_t$ is a real-valued random variable (RV)
• let $\{X_t, t \in \mathbb{Z}\}$ denote a stochastic process, i.e., collection of RVs indexed by ‘time’ $t$ (here $\mathbb{Z}$ denotes the set of all integers)
• use $j$th level equivalent MODWT filter $\{\tilde{h}_{j,l}\}$ on $\{X_t\}$ to create a new stochastic process:

$$
\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t \in \mathbb{Z},
$$

which should be contrasted with

$$
\tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N}, \quad t = 0, 1, \ldots, N - 1
$$
Theoretical Wavelet Variance: II

- if $Y$ is any RV, let $E\{Y\}$ denote its expectation
- let $\text{var}\ \{Y\}$ denote its variance: $\text{var}\ \{Y\} \equiv E\{(Y - E\{Y\})^2\}$
- definition of time dependent wavelet variance:
  $$\nu_{X,t}^2(\tau_j) \equiv \text{var}\ \{\overline{W}_{j,t}\},$$
  with conditions on $X_t$ so that $\text{var}\ \{\overline{W}_{j,t}\}$ exists and is finite
- $\nu_{X,t}^2(\tau_j)$ depends on $\tau_j$ and $t$
- will focus on time independent wavelet variance
  $$\nu_X^2(\tau_j) \equiv \text{var}\ \{\overline{W}_{j,t}\}$$
  (can adapt theory to handle time varying situation)
- $\nu_X^2(\tau_j)$ well-defined for stationary processes and certain related processes, so let’s review concept of stationarity
Definition of a Stationary Process

• if $U$ and $V$ are two RVs, denote their covariance by

$$\text{cov} \{U, V\} = E\{(U - E\{U\})(V - E\{V\})\}$$

• stochastic process $X_t$ called stationary if
  - $E\{X_t\} = \mu_X$ for all $t$, i.e., constant independent of $t$
  - $\text{cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$, i.e., depends on lag $\tau$, but not $t$

• $s_{X,\tau}$, $\tau \in \mathbb{Z}$, is autocovariance sequence (ACVS)

• $s_{X,0} = \text{cov}\{X_t, X_t\} = \text{var}\{X_t\}$; i.e., variance same for all $t$
Spectral Density Functions: I

- spectral density function (SDF) given by

\[ S_X(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f \tau}, \quad |f| \leq \frac{1}{2} \]

- above requires condition on ACVS such as

\[ \sum_{\tau=-\infty}^{\infty} s_{X,\tau}^2 < \infty \]

(sufficient but not necessary)
Spectral Density Functions: II

- if square summability holds, \( \{s_{X,\tau}\} \leftrightarrow S_X(\cdot) \) says
  \[
  \int_{-1/2}^{1/2} S_X(f) e^{i2\pi f \tau} \, df = s_{X,\tau}, \quad \tau \in \mathbb{Z}
  \]

- setting \( \tau = 0 \) yields fundamental result:
  \[
  \int_{-1/2}^{1/2} S_X(f) \, df = s_{X,0} = \text{var} \{X_t\};
  \]
  i.e., SDF decomposes \( \text{var} \{X_t\} \) across frequencies \( f \)

- interpretation: \( S_X(f) \Delta f \) is the contribution to \( \text{var} \{X_t\} \) due to frequencies in a small interval of width \( \Delta f \) centered at \( f \)
White Noise Process: I

• simplest example of a stationary process is ‘white noise’
• process $X_t$ said to be white noise if
  – it has a constant mean $E\{X_t\} = \mu_X$
  – it has a constant variance $\text{var} \{X_t\} = \sigma_X^2$
  – $\text{cov} \{X_t, X_{t+\tau}\} = 0$ for all $t$ and nonzero $\tau$; i.e., distinct RVs in the process are uncorrelated
• ACVS and SDF for white noise take very simple forms:

$$s_{X,\tau} = \text{cov} \{X_t, X_{t+\tau}\} = \begin{cases} \sigma_X^2, & \tau = 0; \\ 0, & \text{otherwise}. \end{cases}$$

$$S_X(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f \tau} = s_{X,0}$$
White Noise Process: II

- ACVS (left-hand plot), SDF (middle) and a portion of length $N = 64$ of one realization (right) for a white noise process with $\mu_X = 0$ and $\sigma^2_X = 1.5$

- since $S_X(f) = 1.5$ for all $f$, contribution $S_X(f) \Delta f$ to $\sigma^2_X$ is the same for all frequencies
Wavelet Variance for Stationary Processes

- for stationary processes, wavelet variance decomposes \( \text{var} \{ X_t \} \):
  \[
  \sum_{j=1}^{\infty} \nu^2_{X}(\tau_j) = \text{var} \{ X_t \}
  \]
  (above result similar to one for sample variance)

- \( \nu^2_{X}(\tau_j) \) is thus contribution to \( \text{var} \{ X_t \} \) due to scale \( \tau_j \)

- note: \( \nu_{X}(\tau_j) \) has same units as \( X_t \), which is important for interpretability
Wavelet Variance for White Noise Process: I

- for a white noise process, can conclude from Exer. [8.1] that
  \[ \nu_X^2(\tau_j) \propto \tau_j^{-1} \]

- note that
  \[ \log (\nu_X^2(\tau_j)) \propto -\log (\tau_j), \]
  so plot of \(\log (\nu_X^2(\tau_j))\) vs. \(\log (\tau_j)\) is linear with a slope of \(-1\)
Wavelet Variance for White Noise Process: II

- $\nu^2_X(\tau_j)$ versus $\tau_j$ for $j = 1, \ldots, 8$ (left-hand plot), along with sample of length $N = 256$ of Gaussian white noise
- largest contribution to var $\{X_t\}$ is at smallest scale $\tau_1$
- note: later on, we will discuss fractionally differenced (FD) processes that are characterized by a parameter $\delta$; when $\delta = 0$, an FD process is the same as a white noise process
Generalization to Certain Nonstationary Processes

• if wavelet filter is properly chosen, $\nu^2_X(\tau_j)$ well-defined for certain processes with stationary backward differences (increments); these are also known as intrinsically stationary processes

• first order backward difference of $X_t$ is process defined by
  \[ X^{(1)}_t = X_t - X_{t-1} \]

• second order backward difference of $X_t$ is process defined by
  \[ X^{(2)}_t = X^{(1)}_t - X^{(1)}_{t-1} = X_t - 2X_{t-1} + X_{t-2} \]

• $X_t$ said to have $d$th order stationary backward differences if
  \[ Y_t \equiv \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k} \]
  forms a stationary process ($d$ is a nonnegative integer)
Examples of Processes with Stationary Increments

- 1st column shows, from top to bottom, realizations from
  (a) random walk: $X_t = \sum_{u=1}^{t} \epsilon_u$, & $\epsilon_t$ is zero mean white noise
  (b) like (a), but now $\epsilon_t$ has mean of $-0.2$
  (c) random run: $X_t = \sum_{u=1}^{t} Y_u$, where $Y_t$ is a random walk

- 2nd & 3rd columns show 1st & 2nd differences $X_t^{(1)}$ and $X_t^{(2)}$
Wavelet Variance for Processes with Stationary Backward Differences: I

- let $\{X_t\}$ be nonstationary with $d$th order stationary differences
- if we use a Daubechies wavelet filter of width $L$ satisfying $L \geq 2d$, then $\nu^2_X(\tau_j)$ is well-defined and finite for all $\tau_j$, but now

$$\sum_{j=1}^{\infty} \nu^2_X(\tau_j) = \infty$$
Wavelet Variance for Random Walk Process: I

- random walk process $X_t = \sum_{u=1}^{t} \epsilon_u$ has first order ($d = 1$) stationary differences since $X_t - X_{t-1} = \epsilon_t$ (i.e., white noise)
- $L \geq 2d$ holds for all wavelets when $d = 1$; for Haar ($L = 2$),
  \[
  \nu^2_X(\tau_j) = \frac{\text{var} \{ \epsilon_t \}}{6} \left( \tau_j + \frac{1}{2\tau_j} \right) \approx \frac{\text{var} \{ \epsilon_t \}}{6} \tau_j,
  \]
  with the approximation becoming better as $\tau_j$ increases
- note that $\nu^2_X(\tau_j)$ increases as $\tau_j$ increases
- $\log (\nu^2_X(\tau_j)) \approx \log (\text{var} \{ \epsilon_t \}/6) + \log (\tau_j)$, which says that a plot of $\log (\nu^2_X(\tau_j))$ vs. $\log (\tau_j)$ is $\approx$ linear with a slope of +1
- as required, also have
  \[
  \sum_{j=1}^{\infty} \nu^2_X(\tau_j) = \frac{\text{var} \{ \epsilon_t \}}{6} \left( 1 + \frac{1}{2} + 2 + \frac{1}{4} + 4 + \frac{1}{8} + \cdots \right) = \infty
  \]
Wavelet Variance for Random Walk Process: II

$\delta = 1$
slope $\approx 1$

- $\nu^2_X(\tau_j)$ versus $\tau_j$ for $j = 1, \ldots, 8$ (left-hand plot), along with sample of length $N = 256$ of a Gaussian random walk process
- smallest contribution to var $\{X_t\}$ is at smallest scale $\tau_1$
- note: a fractionally differenced process with parameter $\delta = 1$ is the same as a random walk process
Wavelet Variance for Processes with Stationary Backward Differences: II

- to see why $\nu_X^2(\tau_j)$ is well-defined and finite if $L \geq 2d$, need basic result from filtering theory: if $\{Y_t\}$ stationary with SDF $S_Y(\cdot)$, then

$$Z_t \equiv \sum_{m=0}^{M-1} a_m Y_{t-m}$$

is also a stationary process, and its SDF is

$$S_Z(f) = A(f) S_Y(f), \text{ where } A(f) \equiv \left| \sum_{m=0}^{M-1} a_m e^{-i2\pi f m} \right|^2,$$

from which it follows that its variance is

$$\text{var} \{Z_t\} = \int_{-1/2}^{1/2} S_Z(f) \, df = \int_{-1/2}^{1/2} A(f) S_Y(f) \, df.$$
Wavelet Variance for Processes with Stationary Backward Differences: III

- example: first backward difference $Y_t^{(1)} = Y_t - Y_{t-1}$, i.e.,
  \[ \{Y_t\} \rightarrow \{1, -1\} \rightarrow \{Y_t^{(1)}\} \]

- here $a_0 = 1$, $a_1 = -1$ and $a_m = 0$ otherwise, yielding
  \[ A(f) = 4 \sin^2(\pi f) \equiv \mathcal{D}(f) \]

(proof of the above is Exer. [105b])
Wavelet Variance for Processes with Stationary Backward Differences: IV

- consider $\nu_{\mathcal{X}}^2(\tau_1)$ (Exer. [304] generalizes result for $\tau_j, j \geq 2$)
- by definition, $\nu_{\mathcal{X}}^2(\tau_1) \equiv \text{var}\{\overline{W}_{1,t}\}$, with $\overline{W}_{1,t} \equiv \sum_{l=0}^{L-1} \tilde{h}_l X_{t-l}$
- because $\tilde{h}_l = h_l/\sqrt{2}$, have

$$\tilde{\mathcal{H}}^{(D)}_1(f) = \frac{1}{2} \mathcal{H}^{(D)}(f) = \sin^L(\pi f) \sum_{l=0}^{L-1} \left( \frac{L}{2} - 1 + l \right) \cos^{2l}(\pi f)$$

$$= D^{L/2}_1(f) \tilde{\mathcal{A}}_L(f)$$

where, as before, $D(f) = 4 \sin^2(\pi f)$ and

$$\tilde{\mathcal{A}}_L(f) \equiv \frac{1}{2L} \sum_{l=0}^{L-1} \left( \frac{L}{2} - 1 + l \right) \cos^{2l}(\pi f)$$
Wavelet Variance for Processes with Stationary Backward Differences: V

• interpret $\tilde{\mathcal{H}}_1^{(D)}(f) = D^\frac{L}{2}(f)\tilde{A}_L(f)$ as the squared gain function for filter cascade consisting of three parts

• first part of cascade consists of a cascade of $d$ first differences:

$$
\{X_t\} \rightarrow \{1, -1\} \rightarrow \cdots \rightarrow \{1, -1\} \rightarrow \{Y_t\}
$$

$d$ of these

where $\{Y_t\}$ is stationary with SDF $S_Y(\cdot)$

• if $\frac{L}{2} > d$, second part uses $\frac{L}{2} - d$ first differences:

$$
\{Y_t\} \rightarrow \{1, -1\} \rightarrow \cdots \rightarrow \{1, -1\} \rightarrow \{Z_t\}
$$

$\frac{L}{2} - d$ of these

where $\{Z_t\}$ is stationary with SDF $S_Z(f) = D^{\frac{L}{2} - d}(f)S_Y(f)$
Wavelet Variance for Processes with Stationary Backward Differences: VI

• third part uses averaging filter embedded within Daubechies wavelet filter:

\[
\{Z_t\} \rightarrow \hat{\mathcal{A}}_L(\cdot) \rightarrow \{\overline{W}_{1,t}\},
\]

where \(\{\overline{W}_{1,t}\}\) is stationary with SDF given by

\[
S_1(f) \equiv \hat{\mathcal{A}}_L(f)S_Z(f) = \mathcal{D}^{L-d}(f)\hat{\mathcal{A}}_L(f)S_Y(f) = \mathcal{D}^{L-d}(f)\hat{\mathcal{A}}_L(f)\mathcal{D}^d(f)S_X(f) = \mathcal{H}_1^{(D)}(f)S_X(f)
\]

if we define an SDF for the nonstationary process \(\{X_t\}\) via

\[
S_X(f) \equiv \frac{S_Y(f)}{\mathcal{D}^d(f)} = \frac{S_Y(f)}{[4\sin^2(\pi f)]^d}
\]

(Yaglom, 1958)
for general $\tau_j$, can claim that, if $\{X_t\}$ has stationary increments of order $d$ and if we use a Daubechies MODWT wavelet filter $\{\tilde{h}_l\}$ of width $L \geq 2d$, the fact that the resulting process $\{\overline{W}_{j,t}\}$ is stationary with variance $\nu_X^2(\tau_j)$ says that

$$\nu_X^2(\tau_j) = \int_{-1/2}^{1/2} \hat{\mathcal{H}}_j^{(D)}(f) S_X(f) \, df,$$

where $\hat{\mathcal{H}}_j^{(D)}(\cdot)$ is the squared gain function for the $j$th level equivalent filter $\{\tilde{h}_{j,l}\}$.
Fractionally Differenced (FD) Processes: I

- can create a continuum of processes that ‘interpolate’ between white noise and random walks using notion of ‘fractional differencing’ (Granger and Joyeux, 1980; Hosking, 1981)
- FD(δ) process is determined by 2 parameters δ and \( \sigma^2_\epsilon \), where \(-\infty < \delta < \infty\) and \( \sigma^2_\epsilon > 0 \) (\( \sigma^2_\epsilon \) is less important than δ)
- if \( \{X_t\} \) is an FD(δ) process, its SDF is given by
  \[
  S_X(f) = \frac{\sigma^2_\epsilon}{D^\delta(f)} = \frac{\sigma^2_\epsilon}{[4\sin^2(\pi f)]^\delta}
  \]
- if \( \delta < 1/2 \), FD process \( \{X_t\} \) is stationary, and, in particular,
  - reduces to white noise if \( \delta = 0 \)
  - has ‘long memory’ or ‘long range dependence’ if \( \delta > 0 \)
  - is ‘antipersistent’ if \( \delta < 0 \) (i.e., cov \( \{X_t, X_{t+1}\} < 0 \)
Fractionally Differenced (FD) Processes: II

- if $\delta \geq 1/2$, FD process $\{X_t\}$ is nonstationary with $d$th order stationary backward differences $\{Y_t\}$
  - here $d = \lfloor \delta + 1/2 \rfloor$, where $\lfloor x \rfloor$ is integer part of $x$
  - $\{Y_t\}$ is stationary FD$(\delta - d)$ process
- if $\delta = 1$, FD process is the same as a random walk process
- using $\sin(x) \approx x$ for small $x$, can claim that, at low frequencies,
  \[
  S_X(f) = \frac{\sigma_\epsilon^2}{[4\sin^2(\pi f)]^\delta} \approx \frac{\sigma_\epsilon^2}{(2\pi f)^{2\delta}}
  \]
  (approximation quite good for $f \in (0, 0.1]$)
- right-hand side describes SDF for a ‘power law’ process with exponent $-2\delta$
Fractionally Differenced (FD) Processes: III

- except possibly for two or three smallest scales, have

\[
\nu_{X}^2(\tau_j) = \int_{-1/2}^{1/2} \tilde{\mathcal{H}}_j^{(D)}(f) S_X(f) \, df
\]

\[
\approx 2 \int_{1/2j+1}^{1/2j} \sigma_{\varepsilon}^2 \frac{1}{[4 \sin^2(\pi f)]^\delta} \, df
\]

\[
\approx \frac{2\sigma_{\varepsilon}^2}{(2\pi)^{2\delta}} \int_{1/2j+1}^{1/2j} \frac{1}{f^{2\delta}} \, df = C\tau_j^{2\delta-1}
\]

- thus \( \log(\nu_{X}^2(\tau_j)) \approx \log(C) + (2\delta - 1) \log(\tau_j) \), so a log/log plot of \( \nu_{X}^2(\tau_j) \) vs. \( \tau_j \) looks approximately linear with slope \( 2\delta - 1 \) for \( \tau_j \) large enough
LA(8) Wavelet Variance for 2 FD Processes

- left-hand column: $\nu_X^2(\tau_j)$ versus $\tau_j$ based upon LA(8) wavelet
- right-hand: realization of length $N = 256$ from each FD process
- see overhead 16 for $\delta = 0$ (white noise), which has slope $= -1$
LA(8) Wavelet Variance for 2 More FD Processes

\[ \delta = \frac{5}{6} \]

\[ \delta = 1 \]

- \( \delta = \frac{5}{6} \) is Kolmogorov turbulence; \( \delta = 1 \) is random walk
- note: positive slope indicates nonstationarity, while negative slope indicates stationarity
Expected Value of Wavelet Coefficients

• in preparation for considering problem of estimating $\nu^2_X(\tau_j)$ given an observed time series, let us consider $E\{\overline{W}_{j,t}\}$

• if $\{X_t\}$ is nonstationary but has $d$th order stationary increments, let $\{Y_t\}$ be the stationary process obtained by differencing $\{X_t\}$ a total of $d$ times; if $\{X_t\}$ is stationary, let $Y_t = X_t$

• Exer. [305]: with $\mu_Y \equiv E\{Y_t\}$, have
  - $E\{\overline{W}_{j,t}\} = 0$ if either (i) $L > 2d$ or (ii) $L = 2d$ and $\mu_Y = 0$
  - $E\{\overline{W}_{j,t}\} \neq 0$ if $\mu_Y \neq 0$ and $L = 2d$

• thus have $E\{\overline{W}_{j,t}\} = 0$ if $L$ is picked large enough ($L > 2d$ is sufficient, but might not be necessary)

• as the argument that follows shows, highly desirable to have $E\{\overline{W}_{j,t}\} = 0$ in order to ease the job of estimating $\nu^2_X(\tau_j)$
Estimation of a Process Variance: I

- suppose \( \{U_t\} \) is a stationary process with mean \( \mu_U = E\{U_t\} \) and unknown variance \( \sigma_U^2 = E\{(U_t - \mu_U)^2\} \)
- can be difficult to estimate \( \sigma_U^2 \) for a stationary process
- to understand why, assume first that \( \mu_U \) is known
- when this is the case, can estimate \( \sigma_U^2 \) using

\[
\tilde{\sigma}_U^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (U_t - \mu_U)^2
\]

- estimator above is unbiased: \( E\{\tilde{\sigma}_U^2\} = \sigma_U^2 \)
Estimation of a Process Variance: II

• if $\mu_U$ is unknown (more common case), can estimate $\sigma^2_U$ using

$$\hat{\sigma}^2_U \equiv \frac{1}{N} \sum_{t=0}^{N-1} (U_t - \bar{U})^2,$$

where $\bar{U} \equiv \frac{1}{N} \sum_{t=0}^{N-1} U_t$

• can argue that $E\{\hat{\sigma}^2_U\} = \sigma^2_U - \text{var} \{\bar{U}\}$

• implies $0 \leq E\{\hat{\sigma}^2_U\} \leq \sigma^2_U$ because var $\{\bar{U}\} \geq 0$

• $E\{\hat{\sigma}^2_U\} \to \sigma^2_U$ as $N \to \infty$ if SDF exists . . . but, for any

$$\epsilon > 0 \text{ (say, 0.00} \cdots 01) \text{ and sample size } N \text{ (say, } N = 10^{10^{10}}\text{)},$$

there is some FD($\delta$) process $\{U_t\}$ with $\delta$ close to 1/2 such that

$$E\{\hat{\sigma}^2_U\} < \epsilon \cdot \sigma^2_U;$$

i.e., in general, $\hat{\sigma}^2_U$ can be badly biased even for very large $N$
Estimation of a Process Variance: III

- example: realization of FD(0.4) process ($\sigma_U^2 = 1 \& N = 1000$)

- using $\mu_U = 0$ (lower horizontal line), obtain $\hat{\sigma}_U^2 \doteq 0.99$

- using $\overline{U} \doteq 0.53$ (upper line), obtain $\hat{\sigma}_U^2 \doteq 0.71$

- note that this is comparable to $E\{\hat{\sigma}_U^2\} \doteq 0.75$

- for this particular example, we would need $N \geq 10^{10}$ to get $\sigma_U^2 - E\{\hat{\sigma}_U^2\} \leq 0.01$, i.e., to reduce the bias so that it is no more than 1% of true variance $\sigma_U^2 = 1$
Estimation of a Process Variance: IV

• conclusion: $\hat{\sigma}_U^2$ can have substantial bias if $\mu_U$ is unknown (can patch up by estimating $\delta$, but must make use of model)

• if $\{X_t\}$ stationary with mean $\mu_X$, then, because $\sum_l \tilde{h}_{j,l} = 0$,

$$E\{\overline{W}_{j,t}\} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} E\{X_{t-l}\} = \mu_X \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0$$

• because $E\{\overline{W}_{j,t}\}$ is known, we can form an unbiased estimator of $\text{var} \\overline{W}_{j,t} = \nu^2_X(\tau_j)$

• more generally, if $\{X_t\}$ is nonstationary with stationary increments of order $d$, we can ensure $E\{\overline{W}_{j,t}\} = 0$ if we pick the filter width $L$ such that $L > 2d$ (in some cases, we might be able to get away with just $L = 2d$)
Wavelet Variance for Processes with Stationary Backward Differences: VIII

• conclusions: $\nu_X^2(\tau_j)$ well-defined for $\{X_t\}$ that is
  - stationary: any $L$ will do and $E\{\overline{W}_{j,t}\} = 0$
  - nonstationary with $d$th order stationary increments: need at least $L \geq 2d$, but might need $L > 2d$ to get $E\{\overline{W}_{j,t}\} = 0$

• if $\{X_t\}$ is stationary, then

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \text{var} \{X_t\} < \infty$$

(recall that each RV in a stationary process must have the same finite variance)
Wavelet Variance for Processes with Stationary Backward Differences: IX

• if \( \{X_t\} \) is nonstationary, then
\[
\sum_{j=1}^{\infty} \nu^2_X(\tau_j) = \infty
\]

• with a suitable construction, we can take the variance of a nonstationary process with \( d \)th order stationary increments to be \( \infty \)

• using this construction, we have
\[
\sum_{j=1}^{\infty} \nu^2_X(\tau_j) = \text{var} \ \{X_t\}
\]

for both the stationary and nonstationary cases

WMTSA: 299–301, 305
Background on Gaussian Random Variables

- \( \mathcal{N}(\mu, \sigma^2) \) denotes a Gaussian (normal) RV with mean \( \mu \) and variance \( \sigma^2 \)
- will write
  \[
  X \overset{d}{=} \mathcal{N}(\mu, \sigma^2)
  \]
to mean ‘RV \( X \) has same distribution as Gaussian RV’
- RV \( \mathcal{N}(0, 1) \) often written as \( Z \) (called standard Gaussian or standard normal)
- let \( \Phi(\cdot) \) be Gaussian cumulative distribution function
  \[
  \Phi(z) \equiv P[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx
  \]
- inverse \( \Phi^{-1}(\cdot) \) of \( \Phi(\cdot) \) is such that \( P[Z \leq \Phi^{-1}(p)] = p \)
- \( \Phi^{-1}(p) \) called \( p \times 100\% \) percentage point
Background on Chi-Square Random Variables

- $X$ said to be a chi-square RV with $\eta$ degrees of freedom if its probability density function (PDF) is given by
  \[ f_X(x; \eta) = \frac{1}{2^{\eta/2}\Gamma(\eta/2)}x^{(\eta/2)-1}e^{-x/2}, \quad x \geq 0, \ \eta > 0 \]

- $\chi^2_\eta$ denotes RV with above PDF

- 3 important facts: $E\{\chi^2_\eta\} = \eta$; $\text{var} \{\chi^2_\eta\} = 2\eta$; and, if $\eta$ is a positive integer and if $Z_1, \ldots, Z_\eta$ are independent $\mathcal{N}(0,1)$ RVs, then
  \[ Z_1^2 + \cdots + Z_\eta^2 \xrightarrow{d} \chi^2_\eta \]

- let $Q_\eta(p)$ denote the $p$th percentage point for the RV $\chi^2_\eta$:
  \[ P[\chi^2_\eta \leq Q_\eta(p)] = p \]
Unbiased Estimator of Wavelet Variance: I

- given a realization of $X_0, X_1, \ldots, X_{N-1}$ from a process with $d$th order stationary differences, want to estimate $\nu^2_X(\tau_j)$
- for wavelet filter such that $L \geq 2d$ and $E\{\overline{W}_{j,t}\} = 0$, have
  $$\nu^2_X(\tau_j) = \text{var}\{\overline{W}_{j,t}\} = E\{\overline{W}^2_{j,t}\}$$
- can base estimator on squares of
  $$\tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N}, \quad t = 0, 1, \ldots, N - 1$$
- recall that
  $$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t \in \mathbb{Z}$$
Unbiased Estimator of Wavelet Variance: II

• comparing

\[ \hat{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N} \]

with \[ \overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l} \]

says that \( \hat{W}_{j,t} = \overline{W}_{j,t} \) if ‘mod \( N \)’ not needed; this happens when \( L_j - 1 \leq t < N \) (recall that \( L_j = (2^j - 1)(L - 1) + 1 \))

• if \( N - L_j \geq 0 \), unbiased estimator of \( \nu^2_X(\tau_j) \) is

\[ \hat{\nu}^2_X(\tau_j) = \frac{1}{N - L_j + 1} \sum_{t=L_j-1}^{N-1} \hat{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2, \]

where \( M_j \equiv N - L_j + 1 \)
Statistical Properties of $\hat{\nu}_X^2(\tau_j)$

- assume that $\{\overline{W}_{j,t}\}$ is Gaussian stationary process with mean zero and ACVS $\{s_{j,\tau}\}$
- suppose $\{s_{j,\tau}\}$ is such that

$$A_j \equiv \sum_{\tau=-\infty}^{\infty} s_{j,\tau}^2 < \infty$$

(if $A_j = \infty$, can make it finite usually by just increasing $L$)
- can show that $\hat{\nu}_X^2(\tau_j)$ is asymptotically Gaussian with mean $\nu_X^2(\tau_j)$ and large sample variance $2A_j/M_j$; i.e.,

$$\frac{\hat{\nu}_X^2(\tau_j) - \nu_X^2(\tau_j)}{(2A_j/M_j)^{1/2}} = \frac{M_j^{1/2}(\hat{\nu}_X^2(\tau_j) - \nu_X^2(\tau_j))}{(2A_j)^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

approximately for large $M_j \equiv N - L_j + 1$
Estimation of $A_j$

- in practical applications, need to estimate $A_j = \sum_{\tau} s_{j,\tau}^2$
- can argue that, for large $M_j$, the estimator

$$\hat{A}_j \equiv \frac{\left(\hat{s}_{j,0}^{(p)}\right)^2}{2} + \sum_{\tau=1}^{M_j-1} \left(\hat{s}_{j,\tau}^{(p)}\right)^2,$$

is approximately unbiased, where

$$\hat{s}_{j,\tau}^{(p)} \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1-|\tau|} \widetilde{W}_{j,t} \widetilde{W}_{j,t+|\tau|}, \quad 0 \leq |\tau| \leq M_j - 1$$

- Monte Carlo results: $\hat{A}_j$ reasonably good for $M_j \geq 128$
Confidence Intervals for \( \nu_X^2(\tau_j) \): I

• based upon large sample theory, can form a \( 100(1 - 2p)\% \) confidence interval (CI) for \( \nu_X^2(\tau_j) \):

\[
\left[ \hat{\nu}_X^2(\tau_j) - \Phi^{-1}(1 - p) \frac{\sqrt{2A_j}}{\sqrt{M_j}}, \hat{\nu}_X^2(\tau_j) + \Phi^{-1}(1 - p) \frac{\sqrt{2A_j}}{\sqrt{M_j}} \right];
\]

i.e., random interval traps unknown \( \nu_X^2(\tau_j) \) with probability \( 1 - 2p \)

• if \( A_j \) replaced by \( \hat{A}_j \), approximate \( 100(1 - 2p)\% \) CI

• critique: lower limit of CI can very well be negative even though \( \nu_X^2(\tau_j) \geq 0 \) always

• can avoid this problem by using a \( \chi^2 \) approximation
Confidence Intervals for $\nu^2_X(\tau_j)$: II

• $\chi^2_{\eta}$ useful for approximating distribution of linear combinations of squared Gaussians

• let $U_1, U_2, \ldots, U_K$ be $K$ independent Gaussian RVs with mean 0 & variance $\sigma^2$; then, since $\text{var} \{U_k^2\} = 2\sigma^4$,

$$Q \equiv \sum_{k=1}^{K} \lambda_k U_k^2$$

has $E\{Q\} = \sigma^2 \sum_{k=1}^{K} \lambda_k$ & $\text{var} \{Q\} = 2\sigma^4 \sum_{k=1}^{K} \lambda_k^2$

• take distribution of $Q$ to be that of the RV $a\chi^2_{a\eta}$, where $a$ and equivalent degrees of freedom (EDOF) $\eta$ are to be determined

• because $E\{\chi^2_{\eta}\} = \eta$ and $\text{var} \{\chi^2_{\eta}\} = 2\eta$, we have $E\{a\chi^2_{a\eta}\} = a\eta$ and $\text{var} \{a\chi^2_{a\eta}\} = 2a^2\eta$

• can equate $E\{Q\}$ & $\text{var} \{Q\}$ to $a\eta$ & $2a^2\eta$ to determine $a$ & $\eta$
Confidence Intervals for $\nu^2_X(\tau_j)$: III

- obtain

$$E\{Q\} = a\eta = \sigma^2 \sum_{k=1}^{K} \lambda_k$$ and

$$\text{var}\{Q\} = 2a^2\eta = 2\sigma^4 \sum_{k=1}^{K} \lambda_k^2,$$

which, when combined, yield

$$\eta = \frac{2(E\{Q\})^2}{\text{var}\{Q\}} = \frac{(\sum_{k=1}^{K} \lambda_k)^2}{\sum_{k=1}^{K} \lambda_k^2}$$ and

$$a = \sigma^2 \frac{\sum_{k=1}^{K} \lambda_k^2}{\sum_{k=1}^{K} \lambda_k}$$

- can also use to approximate sums of correlated squared Gaussians with zero means, e.g.,

$$\hat{\nu}^2_X(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} W_{j,t}^2$$

- can determine $\eta$ based upon

$$E\{\hat{\nu}^2_X(\tau_j)\} = \nu^2_X(\tau_j)$$ and an approximation for

$$\text{var}\{\hat{\nu}^2_X(\tau_j)\}$$
Three Ways to Set $\eta$: I

1. use large sample theory with appropriate estimates:

$$\eta = \frac{2\left(E\{\hat{\nu}_X^2(\tau_j)\}\right)^2}{\text{var} \left\{ \hat{\nu}_X^2(\tau_j) \right\}} \approx \frac{2\nu_X^4(\tau_j)}{2A_j/M_j}$$

suggests $\hat{\eta}_1 = \frac{M_j\hat{\nu}_X^4(\tau_j)}{\hat{A}_j}$

2. assume nominal shape for SDF of $\{X_t\}$: $S_X(f) = hC(f)$, where $C(\cdot)$ is known, but $h$ is not; though questionable, get acceptable CIs using

$$\eta_2 = \frac{2 \left( \sum_{k=1}^{[\frac{(M_j-1)/2}] } C_j(f_k) \right)^2}{\sum_{k=1}^{[\frac{(M_j-1)/2}] } C_j^2(f_k)}$$

\& $C_j(f) \equiv \int_{-1/2}^{1/2} \tilde{H}_j^{(D)}(f)C(f) \, df$

3. make an assumption about the effect of wavelet filter on $\{X_t\}$ to obtain simple (but effective!) approximation

$$\eta_3 = \max\{M_j/2^j, 1\}$$
Three Ways to Set $\eta$: II

- comments on three approaches

1. $\hat{\eta}_1$ requires estimation of $A_j$
   - works well for $M_j \geq 128$ (5% to 10% errors on average)
   - can yield optimistic CIs for smaller $M_j$

2. $\eta_2$ requires specification of shape of $S_X(\cdot)$
   - common practice in, e.g., atomic clock literature

3. $\eta_3$ assumes band-pass approximation
   - default method if $M_j$ small and there is no reasonable guess at shape of $S_X(\cdot)$
Confidence Intervals for $\nu^2_X(\tau_j)$: IV

- after $\eta$ has been determined, can obtain a CI for $\nu^2_X(\tau_j)$
- Exer. [313b]: with prob. $1 - 2p$, the random interval

$$\left[ \frac{\eta\hat{\nu}^2_X(\tau_j)}{Q\eta(1 - p)}, \frac{\eta\hat{\nu}^2_X(\tau_j)}{Q\eta(p)} \right]$$

traps the true unknown $\nu^2_X(\tau_j)$

- lower limit is now nonnegative
- get approximate $100(1 - 2p)\%$ CI for $\nu^2_X(\tau_j)$, with approximation improving as $N \to \infty$, if we use $\hat{\eta}_1$ to estimate $\eta$
- as $N \to \infty$, above CI and Gaussian-based CI converge
Atomic Clock Deviates: I

\[ X_t \]

\[ X_t^{(1)} \]

\[ X_t^{(2)} \]
Atomic Clock Deviates: II

• top plot: errors \( \{X_t\} \) in time kept by atomic clock 571 as compared to time kept at Naval Observatory (measured in microseconds, where 1,000,000 microseconds = 1 second)

• middle: first backward differences \( \{X_t^{(1)}\} \) in nanoseconds (1000 nanoseconds = 1 microsecond)

• bottom: second backward differences \( \{X_t^{(2)}\} \), also in nanoseconds

• if \( \{X_t\} \) nonstationary with \( d \)th order stationary increments, need \( L \geq 2d \), but might need \( L > 2d \) to get \( E\{\overline{W}_{j,t}\} = 0 \)

• Q: what is an appropriate \( L \) here?
Atomic Clock Deviates: III

\[ \tau_j \Delta t \text{ (days)} \]

\[ \text{WMTSA: 319} \]

\[ \text{X-54} \]
Atomic Clock Deviates: IV

• square roots of wavelet variance estimates for atomic clock time errors \( \{X_t\} \) based upon unbiased MODWT estimator with
  – Haar wavelet (x’s in left-hand plot, with linear fit)
  – D(4) wavelet (circles in left- and right-hand plots)
  – D(6) wavelet (pluses in left-hand plot).
• Haar wavelet inappropriate
  – need \( \{X_t^{(1)}\} \) to be a realization of a stationary process with mean 0 (stationarity might be OK, but mean 0 is way off)
  – see Exer. [320b] for explanation of linear appearance
• 95% confidence intervals in the right-hand plot are the square roots of intervals computed using the chi-square approximation with \( \eta \) given by \( \hat{\eta}_1 \) for \( j = 1, \ldots, 6 \) and by \( \eta_3 \) for \( j = 7 \) & 8
Wavelet Variance Analysis of Time Series with Time-Varying Statistical Properties

- each wavelet coefficient $\widehat{W}_{j,t}$ formed using portion of $X_t$
- suppose $X_t$ associated with actual time $t_0 + t \Delta t$
  * $t_0$ is actual time of first observation $X_0$
  * $\Delta t$ is spacing between adjacent observations
- suppose $\tilde{h}_{j,l}$ is least asymmetric Daubechies wavelet
- can associate $\widehat{W}_{j,t}$ with an interval of width $2\tau_j \Delta t$ centered at
  \[ t_0 + (2^j(t + 1) - 1 - |\nu_j^H| \mod N) \Delta t, \]
  where, e.g., $|\nu_j^H| = [7(2^j - 1) + 1]/2$ for LA(8) wavelet
- can thus form ‘localized’ wavelet variance analysis (implicitly assumes stationarity or stationary increments locally)
Subtidal Sea Level Fluctuations: I

- estimated time-dependent LA(8) wavelet variances for physical scale $\tau_2 \Delta t = 1$ day based upon averages over monthly blocks (30.5 days, i.e., 61 data points)
- plot also shows a representative 95% confidence interval based upon a hypothetical wavelet variance estimate of $\frac{1}{2}$ and a chi-square distribution with $\eta = 15.25$
Subtidal Sea Level Fluctuations: II

- estimated LA(8) wavelet variances for physical scales $\tau_j \Delta t = 2^{j-2}$ days, $j = 2, \ldots, 7$, grouped by calendar month
Annual Minima of Nile River

- left-hand plot: annual minima of Nile River
- right: Haar $\hat{\nu}_X^2(\tau_i)$ before (x’s) and after (o’s) year 715.5, with 95% confidence intervals based upon $\chi^2_{\eta_3}$ approximation
Vertical Shear in the Ocean: I

- selected ‘stationary’ portion of vertical shear measurements \( \{X_t\} \) (top plot) and their first backward differences \( \{X_t^{(1)}\} \)
Vertical Shear in the Ocean: II

- unbiased MODWT wavelet variance estimates using the following wavelet filters: Haar (x’s in left-hand plot, through which two regression lines have been fit); D(4) (small circles, right-hand plot); D(6) (pluses, both plots); and LA(8) (big circles, right-hand plot).
D(6) wavelet variance estimates, along with 95% confidence intervals for true wavelet variance with EDOFs determined by, from left to right within each group of 3, \( \hat{\eta}_1 \) (estimated from data), \( \eta_2 \) (using a nominal model for \( S_X(\cdot) \)) and \( \eta_3 = \max\{ M_j/2^j, 1 \} \)
Some Extensions and Ongoing Work

• biased estimators of wavelet variance
• unbiased estimator of wavelet variance for ‘gappy’ time series
• asymptotic theory for non-Gaussian processes satisfying a certain ‘mixing’ condition
• wavelet cross-covariance and cross-correlation
• extension of notion and estimators to random fields
Summary

• wavelet variance gives scale-based analysis of variance

• presented statistical theory for Gaussian processes with stationary increments

• in addition to the applications we have considered, the wavelet variance has been used to analyze
  – genome sequences
  – changes in variance of soil properties
  – canopy gaps in forests
  – accumulation of snow fields in polar regions
  – boundary layer atmospheric turbulence
  – regular and semiregular variable stars