A Tutorial on Stochastic Models and Statistical Analysis for Frequency Stability Measurements

Don Percival

Applied Physics Lab, University of Washington, Seattle

overheads for talk available at

`http://staff.washington.edu/dbp/talks.html`

Introduction

- time scales limited by clock noise
- can model clock noise as stochastic process $\{X_t\}$
  - set of random variables (RVs) indexed by t
  - X_t represents clock noise at time t
  - will concentrate on sampled data, for which will take $t \in \mathbb{Z} \equiv \{\ldots, -1, 0, 1, \ldots\}$ (but sometimes use $t \in \mathbb{Z}^* \equiv \{0, 1, 2, \ldots\}$)
- Q: which stochastic processes are useful models?
- Q: how can we deduce model parameters & other characteristics from observed data?
- will cover the following in this tutorial:
  - stationary processes & closely related processes
  - fractionally differenced & related processes
  - two analysis of variance ('power') techniques
    - $*$ spectral analysis
    - $*$ wavelet analysis
  - parameter estimation via analysis techniques

Stationary Processes: I

- stochastic process $\{X_t\}$ called stationary if
  - $E\{X_t\} = \mu_X$ for all t;
    i.e., a constant that does not depend on t
  - $\text{cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$, all possible t & $t + \tau$;
    i.e., depends on lag τ, 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., process variance is constant for all t
- spectral density function (SDF) given by

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

  note: $S_X(-f) = S_X(f)$ for real-valued processes

Stationary Processes: II

- if $\{X_t\}$ has SDF $S_X(\cdot)$, then
$$\int_{-1/2}^{1/2} S_X(f) e^{i2\pi f\tau}\, df = s_{X,\tau}, \quad \tau \in \mathbb{Z}$$
- 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 var $\{X_t\}$ across frequencies f
- if $\{a_u\}$ is a filter, then (with 'matching condition')
$$Y_t \equiv \sum_{u=-\infty}^{\infty} a_u X_{t-u}$$
is stationary with SDF given by
$$S_Y(f) = \mathcal{A}(f) S_X(f), \text{ where } \mathcal{A}(f) \equiv \left| \sum_{u=-\infty}^{\infty} a_u e^{-i2\pi f u} \right|^2$$
- if $\{a_u\}$ narrow-band of bandwidth Δf about f, i.e.,
$$\mathcal{A}(f') = \begin{cases} \frac{1}{2\Delta f}, & f - \frac{\Delta f}{2} \le |f'| \le f + \frac{\Delta f}{2} \\ 0, & \text{otherwise,} \end{cases}$$
then have following interpretation for $S_X(f)$:
$$\text{var}\,\{Y_t\} = \int_{-1/2}^{1/2} S_Y(f')\, df' = \int_{-1/2}^{1/2} \mathcal{A}(f') S_X(f')\, df' \approx S_X(f)$$

White Noise Process

- simplest stationary process is white noise
- $\{\epsilon_t\}$ is white noise process if
  - $E\{\epsilon_t\} = \mu_\epsilon$ for all t (usually take $\mu_\epsilon = 0$)
  - $\text{var}\,\{\epsilon_t\} = \sigma_\epsilon^2$ for all t
  - $\text{cov}\,\{\epsilon_t, \epsilon_{t'}\} = 0$ for all $t \neq t'$
- white noise thus stationary with ACVS

$$s_{\epsilon,\tau} = \text{cov}\,\{\epsilon_t, \epsilon_{t+\tau}\} = \begin{cases} \sigma_\epsilon^2, & \tau = 0; \\ 0, & \text{otherwise,} \end{cases}$$

  and SDF

$$S_\epsilon(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f\tau} = \sigma_\epsilon^2$$

Backward Differences of White Noise

- consider first order backward difference of white noise:

$$X_t = \epsilon_t - \epsilon_{t-1} = \sum_{u=-\infty}^{\infty} a_u \epsilon_{t-u} \text{ with } a_u \equiv \begin{cases} 1, & u = 0; \\ -1, & u = 1; \\ 0, & \text{otherwise.} \end{cases}$$

- have $S_X(f) = \mathcal{A}(f) S_\epsilon(f) = |2\sin(\pi f)|^2 \sigma_\epsilon^2 \approx |2\pi f|^2 \sigma_\epsilon^2$ at low frequencies (using $\sin(x) \approx x$ for small x)

- let B be backward shift operator: $B\epsilon_t = \epsilon_{t-1}$, $B^2\epsilon_t = \epsilon_{t-2}$, $(1-B)\epsilon_t = \epsilon_t - \epsilon_{t-1}$, etc.

- consider dth order backward difference of white noise:

$$\begin{aligned} X_t = (1-B)^d \epsilon_t &= \sum_{k=0}^{d} \binom{d}{k} (-1)^k \epsilon_{t-k} \\ &= \sum_{k=0}^{d} \frac{d!}{k!(d-k)!} (-1)^k \epsilon_{t-k} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)} (-1)^k \epsilon_{t-k} \end{aligned}$$

  with $\delta \equiv -d$, i.e., $\delta = -1, -2, \ldots$

- SDF given by

$$S_X(f) = \mathcal{A}(f) S_\epsilon(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^{2\delta}}$$

Fractional Differences of White Noise

- for δ not necessary an integer,

  $$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k \epsilon_{t-k} \equiv \sum_{k=0}^{\infty} a_k(\delta)\epsilon_{t-k}$$

  makes sense as long as $\delta < 1/2$

- $\{X_t\}$ stationary fractionally differenced (FD) process
- SDF is as before:

  $$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^{2\delta}}$$

- $\{X_t\}$ said to obey power law at low frequencies if

  $$\lim_{f\to 0} \frac{S_X(f)}{C|f|^\alpha} = 1$$

  for $C > 0$; i.e., $S_X(f) \approx C|f|^\alpha$ at low frequencies

- FD processes obey above with $\alpha = -2\delta$
- note: FD process reduces to white noise when $\delta = 0$

ACVS & PACS for FD Processes

- for $\delta < 1/2$ & $\delta \neq 0, -1, \ldots$, ACVS given by
$$s_{X,\tau} = \sigma_\epsilon^2 \frac{\sin(\pi\delta)\Gamma(1-2\delta)\Gamma(\tau+\delta)}{\pi\Gamma(1+\tau-\delta)};$$
when $\delta = 0, -1, \ldots$, have $s_{X,\tau} = 0$ for $|\tau| > -\delta$ &
$$s_{X,\tau} = \sigma_\epsilon^2 \frac{(-1)^\tau \Gamma(1-2\delta)}{\Gamma(1+\tau-\delta)\Gamma(1-\tau-\delta)}, \quad 0 \leq |\tau| \leq -\delta$$
- for all $\delta < 1/2$, have
$$s_{X,0} = \text{var}\{X_t\} = \sigma_\epsilon^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)},$$
and rest of ACVS can be computed easily via
$$s_{X,\tau} = s_{X,\tau-1} \frac{\tau+\delta-1}{\tau-\delta}, \quad \tau \in \mathbb{Z}^+ \equiv \{1, 2, \ldots\}$$
(for negative lags τ, recall that $s_{X,-\tau} = s_{X,\tau}$).
- for all $\delta < 1/2$, partial autocorrelation sequence (PACS) given by
$$\phi_{t,t} \equiv \frac{\delta}{t-\delta}, \quad t \in \mathbb{Z}^+$$
(useful for constructing best linear predictors)
- FD processes thus have simple and easily computed expressions for SDF, ACVS and PACS

Simulating Stationary FD Processes

- for $-1 \leq \delta < 1/2$, can obtain exact simulations via 'circulant embedding' (Davies–Harte algorithm)
- given $s_{X,0}, \ldots, s_{X,N}$, use discrete Fourier transform (DFT) to compute

$$S_k \equiv \sum_{\tau=0}^{N} s_{X,\tau} e^{-i2\pi f_k \tau} + \sum_{\tau=N+1}^{2N-1} s_{X,2N-\tau} e^{-i2\pi f_k \tau}, \quad k = 0, \ldots, N$$

- given $2N$ independent Gaussian deviates ε_t with mean zero and variance σ_ϵ^2, compute

$$\mathcal{Y}_k \equiv \begin{cases} \varepsilon_0 \sqrt{2NS_0}, & k = 0; \\ (\varepsilon_{2k-1} + i\varepsilon_{2k}) \sqrt{NS_k}, & 1 \leq k < N; \\ \varepsilon_{2N-1} \sqrt{2NS_N}, & k = N; \\ \mathcal{Y}^*_{2N-k}, & N < k \leq 2N-1; \end{cases}$$

(asterisk denotes complex conjugate)

- use inverse DFT to construct the real-valued sequence

$$Y_t = \frac{1}{2N} \sum_{k=0}^{2N-1} \mathcal{Y}_k e^{i2\pi f_k t}, \quad t = 0, \ldots, 2N-1$$

- $Y_0, Y_1, \ldots, Y_{N-1}$ is exact simulation of FD process
- implication: can represent $X_0, X_1, \ldots, X_{N-1}$ as

$$X_t = \sum_{k=0}^{2N-1} c_{t,k}(\delta) \varepsilon_k \text{ rather than } X_t = \sum_{k=0}^{\infty} a_k(\delta) \epsilon_{t-k}$$

Nonstationary FD Processes: I

- suppose $X_t^{(1)}$ is FD process with parameter $\delta^{(s)}$ such that $-1/2 \leq \delta^{(s)} < 1/2$
- define $X_t, t \in \mathbb{Z}^*$, as cumulative sum of $X_t^{(1)}, t \in \mathbb{Z}^*$:

$$X_t \equiv \sum_{l=0}^{t} X_l^{(1)}$$

(for $l < 0$, let $X_t \equiv 0$)

- since, for $t \in \mathbb{Z}^*$,

$$X_t^{(1)} = X_t - X_{t-1} \ \ \& \ \ S_{X^{(1)}}(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta^{(s)}}},$$

filtering theory suggests using relationship

$$S_{X^{(1)}}(f) = |2\sin(\pi f)|^2 S_X(f)$$

to *define* SDF for X_t, i.e.,

$$S_X(f) = \frac{S_{X^{(1)}}(f)}{|2\sin(\pi f)|^2} = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}}$$

with $\delta \equiv \delta^{(s)} + 1$ (Yaglom, 1958)

Nonstationary FD Processes: II

- X_t has stationary 1st order backward differences
- 1 sum defines FD processes for $1/2 \le \delta < 3/2$
- 2 sums define FD processes for $3/2 \le \delta < 5/2$, etc
- X_t has stationary 2nd order backward differences, etc
- if $X_t^{(1)}$ is white noise ($\delta^{(s)} = 0$) so $S_{X^{(1)}}(f) = \sigma_\epsilon^2$, then X_t is random walk ($\delta = 1$) with

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^2} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^2}$$

- if $X_t^{(2)}$ is white noise and if

$$X_t^{(1)} \equiv \sum_{l=0}^{t} X_l^{(2)} \;\; \& \;\; X_t \equiv \sum_{l=0}^{t} X_l^{(1)}, \quad t \in \mathbb{Z}^*,$$

then X_t is random run ($\delta = 2$), and

$$S_X(f) \approx \frac{\sigma_\epsilon^2}{|2\pi f|^4}$$

Summary of FD Processes

- X_t said to be FD process if its SDF is given by

  $$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^{2\delta}} \text{ at low frequencies}$$

- well-defined for any real-valued δ
- FD process obeys power law at low frequencies with exponent $\alpha = -2\delta$
- if $\delta < 1/2$, FD process stationary with
  - ACVS given by

$$s_{X,0} = \sigma_\epsilon^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} \;\&\; s_{X,\tau} = s_{X,\tau-1}\frac{\tau+\delta-1}{\tau-\delta}, \quad \tau \in \mathbb{Z}^+$$

  - PACS given by

$$\phi_{t,t} \equiv \frac{\delta}{t-\delta}, \quad t \in \mathbb{Z}^+$$

- if $\delta \geq 1/2$, FD process nonstationary but its dth order backward difference is stationary FD process with parameter $\delta^{(s)}$, where

  $$d \equiv \lfloor \delta + 1/2 \rfloor \text{ and } \delta^{(s)} \equiv \delta - d$$

  (here $\lfloor x \rfloor$ is largest integer $\leq x$)

Alternatives to FD Processes: I

- fractional Brownian motion (FBM)
  - $B_H(t), 0 \le t < \infty$, has SDF given by

    $$S_{B_H(t)}(f) = \frac{\sigma_X^2 C_H}{|f|^{2H+1}}, \quad -\infty < f < \infty,$$

    where $\sigma_X^2 > 0$, $C_H > 0$ & $0 < H < 1$
    (H called Hurst parameter; C_H depends on H)
  - power law with $-3 < \alpha < -1$
- discrete fractional Brownian motion (DFBM)
  - $B_t, t \in \mathbb{Z}^+$, is DFBM if $B_t = B_H(t)$
  - B_t has SDF given by

    $$S_{B_t}(f) = \sigma_X^2 C_H \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad |f| \le 1/2$$

  - power law at low frequencies with $-3 < \alpha < -1$
  - reduces to random walk if $H = 1/2$

Alternatives to FD Processes: II

- fractional Gaussian noise (FGN)
  - $X_t, t \in \mathbb{Z}^+$, is FGN if $X_t = B_{t+1} - B_t$
  - X_t has SDF given by

$$S_X(f) = 4\,\sigma_X^2 C_H \sin^2(\pi f) \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad |f| \leq 1/2$$

  - power law at low frequencies with $-1 < \alpha < 1$
  - X_t is stationary, with ACVS given by

$$s_{X,\tau} = \frac{\sigma_X^2}{2}\left(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}\right), \quad \tau \in \mathbb{Z},$$

    where $\sigma_X^2 = \mathrm{var}\,\{X_t\}$
  - reduces to white noise if $H = 1/2$

- discrete pure power law (PPL) process
  - SDF given by $S_X(f) = C_S|f|^\alpha, \quad |f| \leq 1/2$
  - if $\alpha > -1$, stationary, but ACVS takes some effort to compute
  - if $\alpha = 0$, reduces to white noise
  - $\alpha \leq -1$, nonstationary but backward differences of certain order are stationary

FD Processes vs. Alternatives

- FD processes cover full range of power laws
  - FBMs, DFBMs and FGNs cover limited range
  - PPL processes also cover full range
- differencing FD process yields another FD process; differencing alternatives yields new type of process
- FD process has simple SDF; if stationary, has simple ACVS & PACS
  - FBM has simple SDF
  - DFBM has complicated SDF
  - FGN has simple ACVS, complicated SDF & PACS
  - PPL has simple SDF, complicated ACVS & PACS
- FD, DFBM, FGN and PPL model sampled noise
  - might be problematic to change sampling rate
  - FBM models unsampled noise
- Fig. 1: comparison of SDFs for FGN, PPL & FD
- Fig. 2: comparison of realizations

Extensions to FD Processes: I

- composite FD processes

$$S_X(f) = \sum_{m=1}^{M} \frac{\sigma_m^2}{|2\sin(\pi f)|^{2\delta_m}};$$

  i.e., linear combinations of independent FD processes

- autoregressive, fractionally integrated, moving average (ARFIMA) processes

  – idea is to replace ϵ_t in

  $$X_t = \sum_{k=0}^{\infty} a_k(\delta)\epsilon_{t-k}$$

  with ARMA process, say,

  $$U_t = \sum_{k=1}^{p} \phi_k U_{t-k} + \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$$

  – yields process with SDF

  $$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \cdot \frac{\left|1 - \sum_{k=1}^{q} \theta_k e^{-i2\pi f k}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_k e^{-i2\pi f k}\right|^2}$$

  – ARMA part can model, e.g., high-frequency structure in noise

Extensions to FD Processes: II

- can define time-varying FD (TVFD) process via

$$X_t = \sum_{k=0}^{\infty} a_k(\delta_t)\epsilon_{t-k}$$

as long as $\delta_t < 1/2$ for all t

  – can use representation

$$X_t = \sum_{k=0}^{2N-1} c_{t,k}(\delta_t)\varepsilon_k, \quad t = 0, 1, \ldots, N-1,$$

  to extend definition to handle arbitrary δ_t

  – Fig. 3: realizations from 4 TVFD processes

  – can also make σ_ϵ^2 time-varying

FD Process Parameter Estimation

- Q: given realization (clock noise) of $X_0, \ldots, X_{N-1}$ from FD process, how can we estimate δ & σ_ϵ^2?
- *many* different estimators have been proposed! (area of active research)
- will concentrate on estimators based on
  - spectral analysis (frequency-based)
  - wavelet analysis (scale-based)
- advantages of spectral and wavelet analysis
  - physically interpretable
  - both are analysis of variance techniques (useful for more than just estimating δ & σ_ϵ^2)
  - can assess need for models more complex than simple FD process (e.g., composite FD process)
  - provide preliminary estimates for more complicated schemes (maximum likelihood estimation)

Estimation via Spectral Analysis

- recall that SDF for FD process given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}}$$

and thus

$$\log\left(S_X(f)\right) = \log\left(\sigma_\epsilon^2\right) - 2\delta \log\left(|2\sin(\pi f)|\right);$$

i.e., plot of $\log\left(S_X(f)\right)$ vs. $\log\left(|2\sin(\pi f)|\right)$ linear with slope of -2δ

- for $0 < f < 1/8$, have $\sin(\pi f) \approx \pi f$, so

$$\log\left(S_X(f)\right) \approx \log\left(\sigma_\epsilon^2\right) - 2\delta \log\left(2\pi f\right);$$

i.e., plot of $\log\left(S_X(f)\right)$ vs. $\log\left(2\pi f\right)$ approximately linear at low frequencies with slope of $-2\delta = \alpha$

- basic scheme
  - estimate $S_X(f)$ via $\hat{S}_X(f)$
  - fit linear model to $\hat{S}_X(f)$ vs. $\log\left(2\pi f\right)$ over low frequencies
  - use estimated slope $\hat{\alpha}$ to estimate δ via $-\hat{\alpha}/2$
  - use estimated intercept to estimate σ_ϵ^2

The Periodogram: I

- basic estimator of $S(f)$ is periodogram:

$$\hat{S}^{(p)}(f) \equiv \frac{1}{N}\left|\sum_{t=0}^{N-1} X_t e^{-i2\pi f t}\right|^2, \qquad |f| \leq 1/2;$$

- represents decomposition of sample variance:

$$\int_{-1/2}^{1/2} \hat{S}^{(p)}(f)\, df = \frac{1}{N}\sum_{t=0}^{N-1} X_t^2$$

- for stationary processes & large N, theory says

$$\hat{S}^{(p)}(f) \stackrel{\mathrm{d}}{=} S(f)\frac{\chi_2^2}{2}, \quad 0 < f < 1/2,$$

  approximately, implying that

  - $E\{\hat{S}^{(p)}(f)\} \approx E\{S(f)\chi_2^2/2\} = S(f)$
  - $\operatorname{var}\{\hat{S}^{(p)}(f)\} \approx \operatorname{var}\{S(f)\chi_2^2/2\} = S^2(f)$

  (in above '$\stackrel{\mathrm{d}}{=}$' means 'equal in distribution,' and χ_2^2 is chi-square RV with 2 degrees of freedom)

- additionally, $\operatorname{cov}\{\hat{S}^{(p)}(f_j), \hat{S}^{(p)}(f_k)\} \approx 0$ for $f_j \equiv j/N$ & $0 < f_j < f_k < 1/2$

The Periodogram: II

- taking log transform yields

$$\log(\hat{S}^{(p)}(f)) \stackrel{\mathrm{d}}{=} \log\left(S(f)\frac{\chi_2^2}{2}\right) = \log(S(f)) + \log\left(\frac{\chi_2^2}{2}\right)$$

- Bartlett & Kendall (1946):

$$E\left\{\log\left(\frac{\chi_\eta^2}{\eta}\right)\right\} = \psi(\eta) - \log(\eta) \;\;\&\;\; \mathrm{var}\left\{\log\left(\frac{\chi_\eta^2}{\eta}\right)\right\} = \psi'(\eta)$$

where $\psi(\cdot)$ & $\psi'(\cdot)$ are di– & trigamma functions

- yields

$$\begin{aligned} E\{\log(\hat{S}^{(p)}(f))\} &= \log(S(f)) + \psi(2) - \log(2) \\ &= \log(S(f)) - \gamma \\ \mathrm{var}\{\log(\hat{S}^{(p)}(f))\} &= \psi'(2) = \pi^2/6 \end{aligned}$$

($\gamma \doteq 0.57721$ is Euler's constant)

The Periodogram: III

- define $Y^{(p)}(f_j) \equiv \log(\hat{S}^{(p)}(f_j)) + \gamma$
- can model $Y^{(p)}(f_j)$ as

$$\begin{aligned} Y^{(p)}(f_j) &\approx \log(S(f_j)) + \epsilon(f_j) \\ &\approx \log(\sigma_\epsilon^2) - 2\delta \log(2\pi f_j) + \epsilon(f_j) \end{aligned}$$

  over low frequencies indexed by $0 < j < J$

- error $\epsilon(f_j)$ in linear regression model such that
  - $E\{\epsilon(f_j)\} = 0$ & $\text{var}\{\epsilon(f_j)\} = \pi^2/6$ (known!)
  - if $\{X_t\}$ Gaussian & $\hat{S}^{(p)}(f_j)$'s uncorrelated, then $\epsilon(f_j)$'s pairwise uncorrelated
  - $\epsilon(f_j) \stackrel{\mathrm{d}}{=} \log(\chi_2^2)$ markedly non-Gaussian
- least squares procedure yields
  - estimates $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$ for δ and σ_ϵ^2
  - estimates of variability in $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$

Multitaper Spectral Estimation: I

- warnings about periodogram:
  - approximations might require N to be *very* large!
  - approximations of questionable validity for nonstationary FD processes
- Fig. 4: periodogram can suffer from 'leakage'
- tapering is technique for alleviating leakage:

$$\hat{S}^{(d)}(f) \equiv \left| \sum_{t=0}^{N-1} a_t X_t e^{-i2\pi f t} \right|^2$$

  - $\{a_t\}$ called data taper (typically bell-shaped curve)
  - $\hat{S}^{(d)}(\cdot)$ called direct spectral estimator
- critique: loses 'information' at end of series (sample size N effectively shortened)
- Thomson (1982): multitapering recovers 'lost info'
- use set of K orthonormal data tapers $\{a_{n,t}\}$:

$$\sum_{t=0}^{N-1} a_{n,t} a_{l,t} = \begin{cases} 1, & \text{if } n = l; \\ 0, & \text{if } n \neq l. \end{cases} \quad 0 \leq n, l \leq K-1$$

Multitaper Spectral Estimation: II

- use $\{a_{n,t}\}$ to form kth direct spectral estimator:

$$\hat{S}_k^{(mt)}(f) \equiv \left| \sum_{t=0}^{N-1} a_{n,t} X_t e^{-i2\pi ft} \right|^2, \quad n = 0, \ldots, K-1$$

- simplest form of multitaper SDF estimator:

$$\hat{S}^{(mt)}(f) \equiv \frac{1}{K} \sum_{n=0}^{K-1} \hat{S}_n^{(mt)}(f)$$

- sinusoidal tapers are one family of multitapers:

$$a_{n,t} = \left\{ \frac{2}{(N+1)} \right\}^{1/2} \sin \left\{ \frac{(n+1)\pi(t+1)}{N+1} \right\}, \quad t = 0, \ldots, N-1$$

(Riedel & Sidorenko, 1995)

- Figs. 5 and 6: example of multitapering
- if $S(\cdot)$ slowly varying around $S(f)$ & if N large,

$$\hat{S}^{(mt)}(f) \stackrel{\mathrm{d}}{=} \frac{S(f)\chi^2_{2K}}{2K}$$

approximately for $0 < f < 1/2$, impling

$$\mathrm{var}\,\{\hat{S}^{(mt)}(f)\} \approx \frac{S^2(f)}{4K^2}\,\mathrm{var}\,\{\chi^2_{2K}\} = \frac{S^2(f)}{K}$$

Multitaper Spectral Estimation: III

- define $Y^{(mt)}(f_j) \equiv \log(\hat{S}^{(mt)}(f_j)) - \psi(K) + \log(K)$
- can model $Y^{(mt)}(f_j)$ as

$$\begin{aligned} Y^{(mt)}(f_j) &\approx \log(S(f_j)) + \eta(f_j) \\ &\approx \log(\sigma_\epsilon^2) - 2\delta \log(2\pi f_j) + \eta(f_j) \end{aligned}$$

  over low frequencies indexed by $0 < j < J$

- error $\eta(f_j)$ in linear regression model such that
  - $E\{\eta(f_j)\} = 0$
  - $\operatorname{var}\{\eta(f_j)\} = \psi'(K)$, a known constant!
  - approximately Gaussian if $K \geq 5$
  - correlated, but with simple structure:

$$\operatorname{cov}\{\eta(f_j), \eta(f_{j+\nu})\} \approx \begin{cases} \psi'(K)\left(1 - \frac{|\nu|}{K+1}\right), & \text{if } |\nu| \leq K+1; \\ 0, & \text{otherwise.} \end{cases}$$

- generalized least squares procedure yields
  - estimates $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$ for δ and σ_ϵ^2
  - estimates of variability in $\hat{\delta}$ and $\hat{\sigma}_\epsilon^2$
- multitaper approach superior to periodogram approach

Discrete Wavelet Transform (DWT)

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ be observed time series (for convenience, assume N integer multiple of 2^{J_0})
- let $\mathcal{W}$ be $N \times N$ orthonormal DWT matrix
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ is vector of DWT coefficients
- orthonormality says $\mathbf{X} = \mathcal{W}^T\mathbf{W}$, so $\mathbf{X} \Leftrightarrow \mathbf{W}$
- can partition $\mathbf{W}$ as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

- $\mathbf{W}_j$ contains $N_j = N/2^j$ wavelet coefficients
  - related to changes of averages at scale $\tau_j = 2^{j-1}$ (τ_j is jth 'dyadic' scale)
  - related to times spaced 2^j units apart
- $\mathbf{V}_{J_0}$ contains $N_{J_0} = N/2^{J_0}$ scaling coefficients
  - related to averages at scale $\lambda_{J_0} = 2^{J_0}$
  - related to times spaced 2^{J_0} units apart

Example: Haar DWT

- Fig. 7: $\mathcal{W}$ for Haar DWT with $N = 16$
  - first 8 rows yield $\mathbf{W}_1 \propto$ *changes* on scale 1
  - next 4 rows yield $\mathbf{W}_2 \propto$ *changes* on scale 2
  - next 2 rows yield $\mathbf{W}_3 \propto$ *changes* on scale 4
  - next to last row yields $\mathbf{W}_4 \propto$ *change* on scale 8
  - last row yields $\mathbf{V}_4 \propto$ *average* on scale 16
- Fig. 8: Haar DWT coefficients for clock 571

DWT in Terms of Filters

- filter $X_0, X_1, \ldots, X_{N-1}$ to obtain

  $$2^{j/2}\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1;$$

  $h_{j,l}$ is jth level wavelet filter (note: circular filtering)
- subsample to obtain wavelet coefficients:

  $$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1,$$

  where $W_{j,t}$ is tth element of $\mathbf{W}_j$
- Figs. 9 & 10: four sets of wavelet filters
- jth wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$ (i.e., scale related to *interval* of frequencies)
- similarly, scaling filters yield $\mathbf{V}_{J_0}$
- Figs. 11 & 12: four sets of scaling filters
- J_0th scaling filter is low-pass with pass-band $[0, \frac{1}{2^{J_0+1}}]$
- as width L of 1st level filters increases,
  - band-pass & low-pass approximations improve
  - # of embedded differencing operations increases (related to # of 'vanishing moments')

DWT-Based Analysis of Variance

- consider 'energy' in time series:

$$\|\mathbf{X}\|^2 = \mathbf{X}^T\mathbf{X} = \sum_{t=0}^{N-1} X_t^2$$

- energy preserved in DWT coefficients:

$$\|\mathbf{W}\|^2 = \|\mathcal{W}\mathbf{X}\|^2 = \mathbf{X}^T\mathcal{W}^T\mathcal{W}\mathbf{X} = \mathbf{X}^T\mathbf{X} = \|\mathbf{X}\|^2$$

- since $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}, \mathbf{V}_{J_0}$ partitions $\mathbf{W}$, have

$$\|\mathbf{W}\|^2 = \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2,$$

  leading to analysis of sample variance:

$$\hat{\sigma}^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1} X_t^2 = \frac{1}{N}\left(\sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2\right)$$

- scale-based decomposition (cf. frequency-based)

Variation: Maximal Overlap DWT

- can eliminate downsampling and use

$$\widetilde{W}_{j,t} \equiv \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

  to define MODWT coefficients $\widetilde{\mathbf{W}}_j$ (& also $\widetilde{\mathbf{V}}_j$)

- unlike DWT, MODWT is not orthonormal
  (in fact MODWT is highly redundant)

- like DWT, can do analysis of variance because

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2$$

- unlike DWT, MODWT works for all samples sizes N
  (i.e., power of 2 assumption is not required)

- Fig. 13: Haar MODWT coefficients for clock 571
  (cf. Fig. 8 with DWT coefficients)

- can use to track time-varying FD process

Definition of Wavelet Variance

- let X_t, $t \in \mathbb{Z}$, be a stochastic process
- run X_t through jth level wavelet filter:

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

- definition of time dependent wavelet variance (also called wavelet spectrum):

$$\nu_{X,t}^2(\tau_j) \equiv \text{var}\,\{\overline{W}_{j,t}\},$$

  assuming $\text{var}\,\{\overline{W}_{j,t}\}$ exists and is finite
- $\nu_{X,t}^2(\tau_j)$ depends on τ_j and t
- will consider time independent wavelet variance:

$$\nu_X^2(\tau_j) \equiv \text{var}\,\{\overline{W}_{j,t}\}$$

  (can be easily adapted to time varying situation)
- rationale for wavelet variance
  - decomposes variance on scale by scale basis
  - useful substitute/complement for SDF

Variance Decomposition

- suppose X_t has SDF $S_X(f)$:

$$\int_{-1/2}^{1/2} S_X(f)\, df = \operatorname{var}\{X_t\};$$

  i.e., decomposes $\operatorname{var}\{X_t\}$ across frequencies f

  – involves uncountably infinite number of f's

  – $S_X(f)\,\Delta f \approx$ contribution to $\operatorname{var}\{X_t\}$ due to f's in interval of length Δf centered at f

  – note: $\operatorname{var}\{X_t\}$ taken to be ∞ for nonstationary processes with stationary backward differences

- wavelet variance analog to fundamental result:

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

  i.e., decomposes $\operatorname{var}\{X_t\}$ across scales τ_j

  – recall DWT/MODWT and sample variance

  – involves countably infinite number of τ_j's

  – $\nu_X^2(\tau_j)$ contribution to $\operatorname{var}\{X_t\}$ due to scale τ_j

  – $\nu_X(\tau_j)$ has same units as X_t (easier to interpret)

Spectrum Substitute/Complement

- because $\tilde{h}_{j,l} \approx$ bandpass over $[1/2^{j+1}, 1/2^j]$,

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

- if $S_X(f)$ 'featureless', info in $\nu_X^2(\tau_j)$ $\Leftrightarrow$ info in $S_X(f)$
- $\nu_X^2(\tau_j)$ more succinct: only 1 value per octave band
- recall SDF for FD process:

$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}} \approx \frac{\sigma_\epsilon^2}{|2\pi f|^{2\delta}}$$

- $(*)$ implies $\nu_X^2(\tau_j) \propto \tau_j^{2\delta-1}$ approximately
- can deduce δ from slope of $\log(\nu_X^2(\tau_j))$ vs. $\log(\tau_j)$
- can estimate δ & σ_ϵ^2 by applying regression analysis to log of estimates of $\nu_X^2(\tau_j)$

Estimation of Wavelet Variance: I

- can base estimator on MODWT of $X_0, X_1, \ldots, X_{N-1}$:

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

(DWT-based estimator possible, but less efficient)

- recall that

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t = 0, \pm 1, \pm 2, \ldots$$

so $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ if mod not needed: $L_j - 1 \leq t < N$

- if $N - L_j \geq 0$, unbiased estimator of $\nu_X^2(\tau_j)$ is

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

where $M_j \equiv N - L_j + 1$

- can also construct biased estimator of $\nu_X^2(\tau_j)$:

$$\tilde{\nu}_X^2(\tau_j) \equiv \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{N} \Big(\sum_{t=0}^{L_j-2} \widetilde{W}_{j,t}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2 \Big)$$

1st sum in parentheses influenced by circularity

Estimation of Wavelet Variance: II

- biased estimator unbiased if $\{X_t\}$ white noise
- biased estimator offers exact analysis of $\hat{\sigma}^2$; unbiased estimator need not
- biased estimator can have better mean square error (Greenhall *et al.*, 1999; need to 'reflect' X_t)

Statistical Properties of $\hat{\nu}_X^2(\tau_j)$

- suppose $\{\overline{W}_{j,t}\}$ Gaussian, mean 0 & SDF $S_j(f)$
- suppose square integrability condition holds:

$$A_j \equiv \int_{-1/2}^{1/2} S_j^2(f)\,df < \infty \;\; \& \;\; S_j(f) > 0$$

  (holds for FD process if L large enough)

- can show $\hat{\nu}_X^2(\tau_j)$ asymptotically normal with mean $\nu_X^2(\tau_j)$ & large sample variance $2A_j/M_j$
- can estimate A_j and use with $\hat{\nu}_X^2(\tau_j)$ to construct confidence interval for $\nu_X^2(\tau_j)$
- example
  - Fig. 14: clock errors $X_t \equiv X_t^{(0)}$ along with differences $X_t^{(i)} \equiv X_t^{(i-1)} - X_{t-1}^{(i-1)}$ for $i = 1, 2$
  - Fig. 15: $\hat{\nu}_X^2(\tau_j)$ for clock errors
  - Fig. 16: $\hat{\nu}_{\overline{Y}}^2(\tau_j)$ for $\overline{Y}_t \propto X_t^{(1)}$
  - Haar $\hat{\nu}_{\overline{Y}}^2(\tau_j)$ related to Allan variance $\sigma_{\overline{Y}}^2(2, \tau_j)$:

$$\nu_{\overline{Y}}^2(\tau_j) = \tfrac{1}{2}\sigma_{\overline{Y}}^2(2, \tau_j)$$

Summary

- fractionally differenced processes are
  - able to cover all power laws
  - easy to work with (SDF, ACVS & PACS simply expressed)
  - extensible to composite, ARFIMA & time-varying processes
- spectral and wavelet analysis can provide
  - estimates of parameters of FD processes
  - decomposition of sample variance across
    - frequencies (in case of spectral analysis)
    - scales (in case of wavelet analysis)
  - complementary analyses
- wavelet analysis has some advantages for clock noise
  - estimates δ & σ_ϵ^2 somewhat better
  - useful with time-varying noise process
  - can deal with polynomial trends (not covered here)
  - results expressed in same units as X_t^2
- a big 'thank you' to conference organizers!