Wavelet Methods for Time Series Analysis

Part VII: Wavelet-Based Bootstrapping

- start with some background on bootstrapping and its rationale
- describe adjustments to the bootstrap that allow it to work with correlated time series
- describe how the decorrelating property of the DWT can be used to develop a wavelet-based bootstrap for certain time series
- describe 'wavestrapping,' an adaptive procedure based upon finding a decorrelating transform from a wavelet packet table

Motivating Question

- let $\mathbf{X} = [X_0, \ldots, X_{N-1}]^T$ be a finite portion of a stationary process with autocovariance sequence(ACVS) $\{s_\tau\}$
- let $\{\rho_\tau\}$ be the corresponding autocorrelation sequence (ACS):
$$\rho_\tau = \frac{s_\tau}{s_0}, \quad \text{where } s_\tau = \text{cov}\{X_t, X_{t+\tau}\} \text{ and } s_0 = \text{var}\{X_t\}$$
- given a time series, we can estimate its ACS at $\tau = 1$ using
$$\hat{\rho}_1 \equiv \frac{\sum_{t=0}^{N-2} X_t X_{t+1}}{\sum_{t=0}^{N-1} X_t^2}$$
under the assumption that $E\{X_t\} = 0$
- Q: given the amount N of data we have, how close can we expect $\hat{\rho}_1$ to be to the true unknown ρ_1?
- i.e., how can we assess the sampling variability in $\hat{\rho}_1$?

Classic Approach – Large Sample Theory: I

- in what follows, let $\mathcal{N}(\mu, \sigma^2)$ denote a Gaussian (normal) random variable (RV) with mean μ and variance σ^2
- if X_t's were independent and identically distributed (IID) so that $\rho_1 = 0$, the distribution of $\hat{\rho}_1$ becomes arbitrarily close to that of an $\mathcal{N}(0, \frac{1}{N})$ RV as $N \to \infty$ (requires suitable conditions)

Classic Approach – Large Sample Theory: II

- more generally, $\hat{\rho}_1$ is close to the distribution of an $\mathcal{N}(\rho_1, \sigma_N^2)$ RV as $N \to \infty$, where
$$\sigma_N^2 \equiv \frac{1}{N} \sum_{\tau=-\infty}^{\infty} \left\{\rho_\tau^2(1 + 2\rho_1^2) + \rho_{\tau+1}\rho_{\tau-1} - 4\rho_1\rho_\tau\rho_{\tau-1}\right\}$$
- in practice, the above result is unappealing because it requires
  - knowledge of the theoretical ACS
  - the ACS to damp down sufficiently fast, which would rule out long memory processes (LMPs)
- while large sample theory has been worked out for $\hat{\rho}_1$ under certain conditions, similar theory for other statistics can be hard to come by

Alternative Approach – Bootstrapping: I

- if X_t's were IID, we could apply 'bootstrapping' to assess the variability in $\hat{\rho}_1$, as follows
- suppose we have the following time series of length $N = 8$, which is a realization of a Gaussian white noise process:

$$\mathbf{x} \doteq [1.9, 2.2, -0.1, 1.0, -0.6, 0.5, -1.3, -0.3]^T,$$

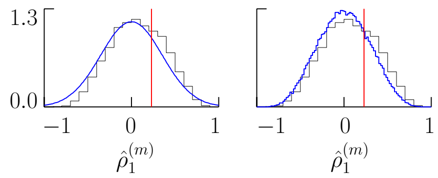
  for which $\hat{\rho}_1 \doteq 0.23$ (for white noise, the true value of ρ_1 is 0)
- generate a new time series $\mathbf{x}^{(1)}$ by randomly sampling from $\mathbf{x}$:

$$\mathbf{x}^{(1)} \doteq [2.2, -0.1, -0.1, 1.0, 1.9, 1.9, -0.6, -0.1]^T,$$

  for which $\hat{\rho}_1^{(1)} \doteq 0.31$ (note: sampling is done with replacement)
- do again to get $\mathbf{x}^{(2)} = [-0.3, 0.5, 1.9, -0.6, -0.3, 0.5, 2.2, 2.2]^T$, for which $\hat{\rho}_1^{(2)} \doteq 0.39$

Alternative Approach – Bootstrapping: II

- repeat a large number of times to get $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \ldots, \hat{\rho}_1^{(M)}$
- plots shows histogram for $\{\hat{\rho}_1^{(m)} : m = 1, \ldots, 10,000\}$, along with probability density function (PDF) for $\mathcal{N}(0, \frac{1}{8})$ (left-hand plot) and an approximation to the true PDF for $\hat{\rho}_1$ (right)

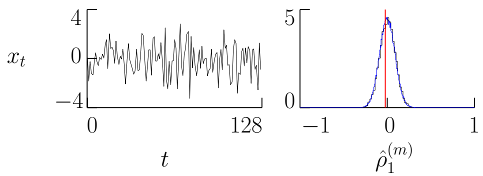


vertical line indicates $\hat{\rho}_1$

- can regard sample distribution of $\{\hat{\rho}_1^{(m)}\}$ as an approximation to the unknown distribution of $\hat{\rho}_1$

Alternative Approach – Bootstrapping: III

- bootstrap approximation to distribution of $\hat{\rho}_1$ gets better as N increases
- consider sample of Gaussian white noise of length $N = 128$, for which $\hat{\rho}_1 \doteq -0.02$

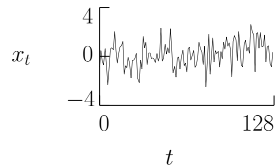


vertical line indicates $\hat{\rho}_1$

- sample distribution of $\{\hat{\rho}_1^{(m)}\}$ agrees quite well with the approximate true PDF

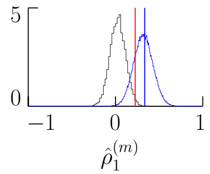
Bootstrapping Correlated Time Series: I

- key assumption: $\mathbf{x}$ was a realization of IID RVs
- if not true (usually the case with time series!), sample distribution of $\{\hat{\rho}_1^{(m)}\}$ can be badly misleading as an approximation to unknown distribution of $\hat{\rho}_1$
- as an example, consider a realization of a fractionally differenced (FD) process with parameter $\delta = \frac{1}{4}$, for which $\hat{\rho}_1 \doteq 0.23$ (for an FD($\frac{1}{4}$) process, $\rho_1 = \frac{1}{3}$)



Bootstrapping Correlated Time Series: II

- use the same procedure as before to get $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \ldots, \hat{\rho}_1^{(M)}$
- plot shows histogram for $\{\hat{\rho}_1^{(m)} : m = 1, \ldots, 10{,}000\}$, along with an approximation to the true PDF for $\hat{\rho}_1$



vertical lines indicate $\hat{\rho}_1$ and ρ_1

- bootstrap approximation gets even worse as N increases
- to correct the problem caused by correlation in time series, can use specialized time or frequency domain bootstrapping *if* ACS damp downs sufficiently fast

Parametric Bootstrapping: I

- one well-known time domain bootstrapping scheme is the parametric (or residual) bootstrap
- suppose we can assume that our time series is a realization of a portion $X_0, \ldots, X_{N-1}$ of a first order autoregressive (AR) process:

$$X_t = \phi_1 X_{t-1} + \epsilon_t,$$

  where $|\phi_1| < 1$ and $\{\epsilon_t\}$ is white noise with zero mean and variance σ_ϵ^2 (this model is widely used in geophysics)
- have $\mathrm{var}\,\{X_t\} = \sigma_\epsilon^2/(1-\phi_1^2)$ and $\rho_\tau = \phi_1^{|\tau|}$ for AR(1) process
- in particular, $\rho_1 = \phi_1$, so can estimate ϕ_1 using $\hat{\phi}_1 \equiv \hat{\rho}_1$

Parametric Bootstrapping: II

- since $\epsilon_t = X_t - \phi_1 X_{t-1}$, can form residuals

$$r_t = X_t - \hat{\phi}_1 X_{t-1}, \quad t = 1, \ldots, N-1,$$

  with the idea that r_t will be a good approximation to ϵ_t (note: there are $N-1$ residuals rather than N)
- let $r_0^{(1)}, r_1^{(1)}, \ldots, r_{N-1}^{(1)}$ be a random sample from $r_1, r_2, \ldots, r_{N-1}$ (as before, sampling is done with replacement)
- let $X_0^{(1)} = r_0^{(1)}/(1-\hat{\phi}_1^2)^{1/2}$ ('stationary initial condition')
- form

$$X_t^{(1)} = \hat{\phi}_1 X_{t-1}^{(1)} + r_t^{(1)}, \quad t = 1, \ldots, N-1,$$

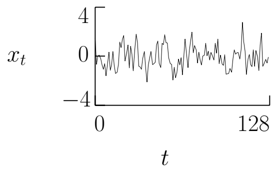
  yielding the bootstrapped time series $X_0^{(1)}, X_1^{(1)}, \ldots, X_{N-1}^{(1)}$

Parametric Bootstrapping: III

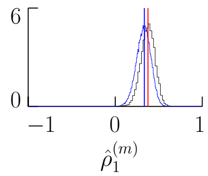
- use $X_0^{(1)}, X_1^{(1)}, \ldots, X_{N-1}^{(1)}$ to compute $\hat{\rho}_1^{(1)}$
- let $r_0^{(2)}, r_1^{(2)}, \ldots, r_{N-1}^{(2)}$ be a second random sample from r_1, r_2, $\ldots$, r_{N-1}
- use these to form a second bootstrapped series $X_0^{(2)}$, $X_1^{(2)}$, $\ldots$, $X_{N-1}^{(2)}$, from which we form $\hat{\rho}_1^{(2)}$
- repeat this procedure M times to get $\hat{\rho}_1^{(1)}, \hat{\rho}_1^{(2)}, \ldots, \hat{\rho}_1^{(M)}$

Parametric Bootstrapping: IV

- as an example, consider a realization of an AR(1) process with $\phi_1 = \rho_1 = \frac{1}{3}$, for which $\hat{\rho}_1 \doteq 0.38$



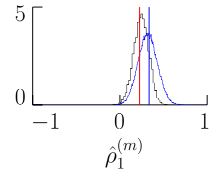
- plot shows histogram for $\{\hat{\rho}_1^{(m)} : m = 1, \ldots, 10,000\}$, along with an approximation to the true PDF for $\hat{\rho}_1$



vertical lines indicate $\hat{\rho}_1$ and ρ_1

Parametric Bootstrapping: V

- important assumption here is that time series is well modeled by AR(1) process
- to see what happens if this assumption fails, reconsider FD($\frac{1}{4}$) realization and treat it as if it were an AR(1) realization
- since $\hat{\rho}_1 \doteq 0.23$, we would set $\hat{\phi}_1 \doteq 0.23$
- plot shows histogram for $\{\hat{\rho}_1^{(m)} : m = 1, \ldots, 10,000\}$, along with an approximation to the true PDF for $\hat{\rho}_1$



vertical lines indicate $\hat{\rho}_1$ and ρ_1

Parametric Bootstrapping: VI

- more generally, can fit pth order process

$$X_t = \sum_{u=1}^{p} \phi_u X_{t-u} + \epsilon_t \text{ and use } r_t = X_t - \sum_{u=1}^{p} \hat{\phi}_u X_{t-u}$$

  to form new series and then $\hat{\rho}_1^{(m)}$
- note that the number of residuals is $N - p$, so best to stick with small values of p
- several variations on the basic scheme, one of which is to use $\tilde{r}_t = r_t - \bar{r}$ rather than r_t, where $\bar{r}$ is the sample mean of the residuals (usually close to zero, but sometimes not)

Block Bootstrapping

- another time domain approach is block bootstrapping, which is nonparametric and has some nice theoretical properties, but a bit trickier to describe and implement

Frequency Domain Bootstrapping

- 'phase scramble' discrete Fourier transform (DFT) $\{\mathcal{X}_k\}$ of data $\{X_t\}$ and apply inverse DFT to create new series:

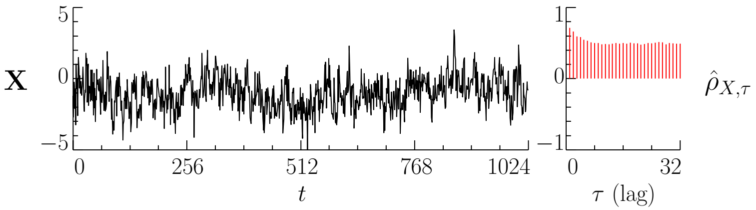
$$\mathcal{X}_k = \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N} = A_k e^{i\theta_k}$$

- periodogram-based bootstrapping: in addition to phase scrambling, evoke large sample result that $|A_k|$'s are approximately uncorrelated with distribution related to a chi-square RV with 2 degrees of freedom
- circulant embedding bootstrapping: form nonparametric estimate of spectral density function and generate realizations using circulant embedding

Rationale for Wavelet Domain Bootstrapping

- time and frequency domain approaches are both problematic for long memory processes
- DWT decorrelates certain time series $\mathbf{X}$, including long memory processes (these are ruled out by time and frequency domain bootstrapping because ACS damps down slowly)
- level J_0 partial DWT maps $\mathbf{X}$ to $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$, with the RVs in the $\mathbf{W}_j$'s being approximately uncorrelated (note: scaling coefficients $\mathbf{V}_{J_0}$ are still highly correlated)

DWT of a Long Memory Process: I

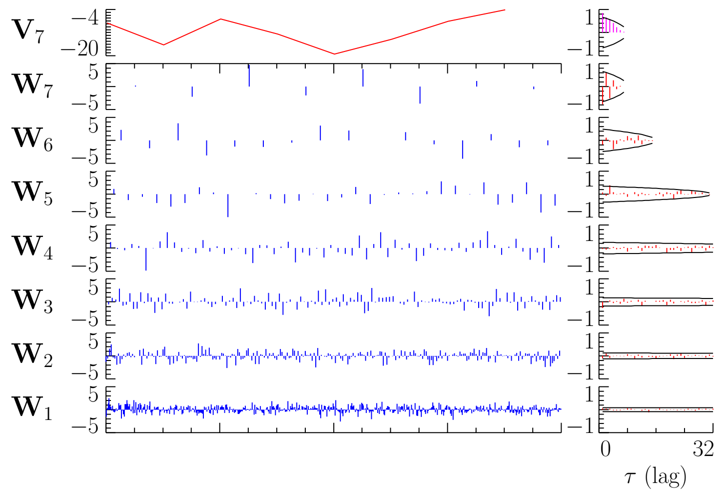


- realization of an FD(0.4) time series $\mathbf{X}$ along with its sample autocorrelation sequence (ACS): for $\tau \geq 0$,

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

- note that ACS dies down slowly

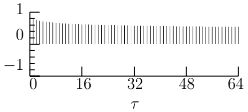
DWT of a Long Memory Process: II



- LA(8) DWT of FD(0.4) series and sample ACSs for each $\mathbf{W}_j$ & $\mathbf{V}_7$, along with 95% confidence intervals for white noise

DWT of a Long Memory Process: III

- second example: ACS for FD(0.45)

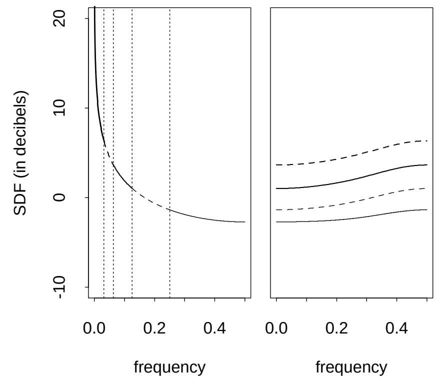


- unit lag autocorrelations for $\mathbf{W}_j$ using the Haar, D(4) and LA(8) wavelet filters (other autocorrelations are very small)

j	Haar	D(4)	LA(8)
1	−0.0626	−0.0797	−0.0767
2	−0.0947	−0.1320	−0.1356
3	−0.1133	−0.1511	−0.1501
4	−0.1211	−0.1559	−0.1535

DWT of a Long Memory Process: IV

- spectral density functions (SDFs) for $\mathbf{X}$ and $\mathbf{W}_j$



  – relatively flat (white noise if perfectly flat), but remaining variation well approximated by SDF for AR(2) process
  – height increases as j increases (variance of $\mathbf{W}_j$ sets height)

DWT of a Long Memory Process: V

- maximum absolute cross-correlations for wavelet coefficients in $\mathbf{W}_j$ and $\mathbf{W}_{j'}$ for $1 \le j < j' \le 4$

	Haar			D(4)			LA(8)		
$j\backslash j'$	2	3	4	2	3	4	2	3	4
1	0.13	0.17	0.14	0.09	0.09	0.04	0.06	0.03	0.00
2		0.17	0.21		0.12	0.11		0.08	0.03
3			0.18			0.13			0.08

Recipe for Wavelet Domain Bootstrapping: I

1. given $\mathbf{X}$ of length $N = 2^J$, compute level $J_0 = J - 2$ partial DWT $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$ (4 coefficients in $\mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$)
2. randomly sample with replacement $N/2^j$ times from $\mathbf{W}_j$ to create bootstrapped vector $\mathbf{W}_j^{(b)}$, $j = 1, \ldots, J_0$
3. do the same for $\mathbf{V}_{J_0}$ to create $\mathbf{V}_{J_0}^{(b)}$ (theory lacking here, but better in computer experiments than using just $\mathbf{V}_{J_0}$)
4. apply inverse transform to $\mathbf{W}_1^{(b)}, \ldots, \mathbf{W}_{J_0}^{(b)}$ and $\mathbf{V}_{J_0}^{(b)}$ to obtain bootstrapped time series $\mathbf{X}^{(b)}$
5. compute unit lag sample autocorrelation $\hat{\rho}_1^{(b)}$

- repeat above many times to build up sample distribution of bootstrapped autocorrelations

Recipe for Wavelet Domain Bootstrapping: II

- computer experiments indicate improvement over block bootstrap for FD processes
- variation: replace $\mathbf{X}$ by series of length $2N$ given by

$$\mathbf{X}_{(c)} \equiv [X_0, X_1, \ldots, X_{N-2}, X_{N-1}, X_{N-1}, X_{N-2}, \ldots, X_1, X_0]^T;$$

i.e., use 'reflection' rather than circular boundary conditions

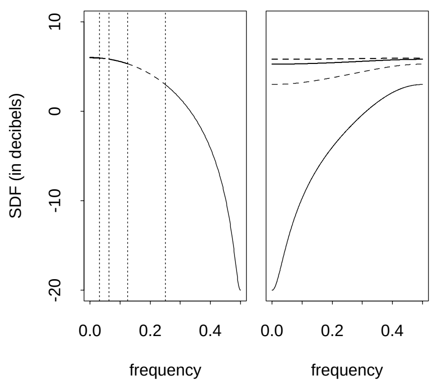
Motivation for 'Wavestrapping': I

- DWT does not adequately decorrelate all time series
- consider first order moving average process (MA(1)):

$$X_t = \epsilon_t + 0.99\epsilon_{t-1}$$

Motivation for 'Wavestrapping': II

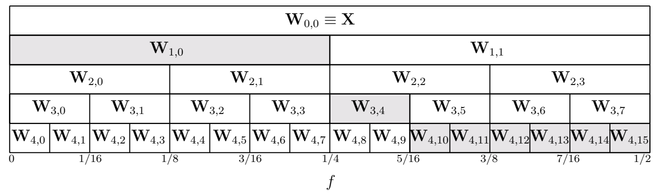
- SDFs for MA(1) process and associated $\mathbf{W}_j$



- note that SDF of $\mathbf{W}_1$ is not approximately flat
- idea: use transform selected from wavelet packet table

Motivation for 'Wavestrapping': III

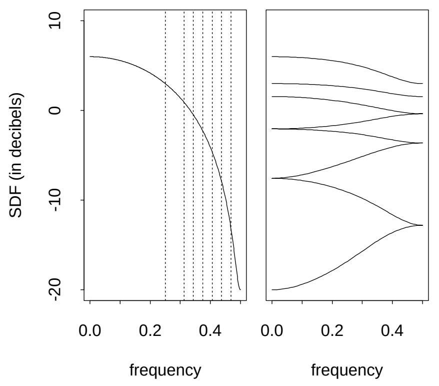
- consider following level $J_0 = 4$ wavelet packet table (WPT):



- shaded boxes identify an orthonormal transform that is a better decorrelator of the MA(1) process than the DWT

Motivation for 'Wavestrapping': IV

- SDFs for MA(1) process and associated $\mathbf{W}_{j,n}$



Motivation for 'Wavestrapping': V

- first 5 of $\mathbf{W}_{j,n}$ SDFs have variations less than 3 dB, but those for $\mathbf{W}_{4,13}$, $\mathbf{W}_{4,14}$ and $\mathbf{W}_{4,15}$ vary by 3.9, 5.3 and 7.2 dB
- increasing depth of WPT to $J_0 = 6$ allows us to replace these by
  - three $j = 5$ level subvectors $\mathbf{W}_{5,26}$, $\mathbf{W}_{5,27}$, $\mathbf{W}_{5,28}$ and
  - six $j = 6$ level subvectors $\mathbf{W}_{6,58}, \ldots, \mathbf{W}_{6,63}$
  - resulting WPT has SDFs that all vary by less than 3 dB
- idea: adaptively select transform by using white noise tests

Recipe for Wavestrapping: I

1. given $\mathbf{X}$ of length 2^J, compute level $J_0 = J - 2$ WPT (enter step 2 with starting values $j = n = 0$ and $\mathbf{W}_{0,0} \equiv \mathbf{X}$)
2. if $j = J_0$, retain $\mathbf{W}_{j,n}$; if $j < J_0$, do white noise test on $\mathbf{W}_{j,n}$
   - portmanteau test on autocorrelation estimates for $\mathbf{W}_{j,n}$
   - cumulative periodogram test

   if fail to reject the null hypothesis, retain $\mathbf{W}_{j,n}$; if reject, discard $\mathbf{W}_{j,n}$ (after transforming it into $\mathbf{W}_{j+1,2n}$ and $\mathbf{W}_{j+1,2n+1}$), and repeat this step twice again (both on $\mathbf{W}_{j+1,2n}$ and $\mathbf{W}_{j+1,2n+1}$)
3. desired adaptively chosen transform consists of all subvectors retained after step 2 applied as many times as needed; randomly sample (with replacement) from each subvector in the transform to create the similarly dimensioned wavestrapped subvectors

Recipe for Wavestrapping: II

4. apply inverse transform to obtain bootstrapped time series $\mathbf{X}^{(b)}$
5. compute unit lag sample autocorrelation $\hat{\rho}_1^{(b)}$

- repeat above many times to build up sample distribution of bootstrapped autocorrelations

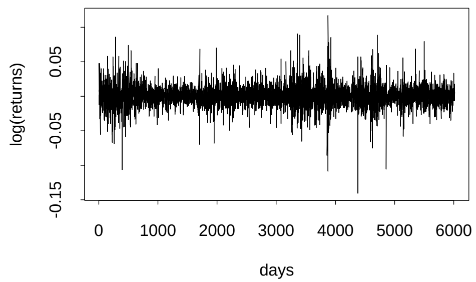
Summary of Computer Experiments - I

Process	Boundary	DWT	Wavestrap Port	Wavestrap Pgrm	Block	True
WN						
$N = 128$	periodic	8.2	8.7	8.8	8.1	8.7
	reflection	8.3	8.6	8.7		
$N = 1024$	periodic	3.1	3.1	3.1	3.0	3.1
	reflection	3.2	3.2	3.1		
AR(1)						
$N = 128$	periodic	5.7	5.2	5.1	5.4	4.8
	reflection	5.5	5.1	5.4		
$N = 1024$	periodic	1.6	1.5	1.5	1.5	1.4
	reflection	1.6	1.5	1.5		
MA(1)						
$N = 128$	periodic	7.1	6.8	6.8	6.5	6.3
	reflection	7.0	6.8	6.6		
$N = 1024$	periodic	2.6	2.4	2.3	2.2	2.2
	reflection	2.6	2.4	2.4		
FD						
$N = 128$	periodic	9.4	8.3	8.5	7.7	10.7
	reflection	9.9	8.8	9.6		
$N = 1024$	periodic	4.4	4.2	4.2	3.4	5.3
	reflection	4.7	4.5	4.7		

Summary of Computer Experiments - II

- standard deviations ($\times 100$) of unit lag sample autocorrelations given by DWT-based bootstrapping, two forms of wavestrapping and block bootstrap, along with true standard deviations
- four models considered are white noise (WN); AR(1) process $X_t = 0.9X_{t-1} + \epsilon_t$; MA(1) process $X_t = \epsilon_t + 0.99\epsilon_{t-1}$; and fractionally differenced (FD) process with $\delta = 0.45$
- wavestrapping does better than block bootstrap (current state of the art) except for the MA(1) process, for which the block bootstrap is ideally suited

Application to BMW Stock Prices - I

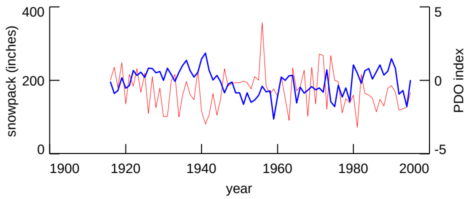


- plot shows log of daily returns on BMW share prices
- has small unit lag sample autocorrelation: $\hat{\rho}_1 \doteq 0.081$.
- large sample theory appropriate for Gaussian white noise gives standard error of $1/\sqrt{N} \doteq 0.013$

Application to BMW Stock Prices - II

- Gaussianity is suspect: data better modeled by t distribution with 3.9 degrees of freedom
- block bootstrap with block sizes 30, 50, 100, 200 and 500 gives standard errors are 0.012, 0.012, 0.014, 0.016 and 0.015
- DWT-based bootstrap and wavestrap give 0.023 & 0.020
- confirms presence of autocorrelation (small, but presumably exploitable by traders)

Applications to Bivariate Climate Time Series - I



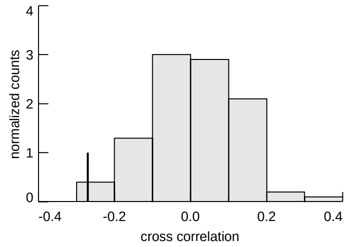
- plot shows Pacific decadal oscillation (PDO) index (thick curve) and March 15th snow depth on Mt. Rainier (thin curve)
- sample cross-correlation is

$$\hat{\rho}_{XY} \equiv \frac{\sum_{t=0}^{N-1}(X_t - \overline{X})(Y_t - \overline{Y})}{\left[\sum_{t=0}^{N-1}(X_t - \overline{X})^2 \sum_{t=0}^{N-1}(Y_t - \overline{Y})^2\right]^{1/2}} \doteq -0.27$$

- Q: given such a short series, is this significantly different from zero?

Applications to Bivariate Climate Time Series - II

- histogram of wavestrapped cross-correlations says ‘yes’



Comments on Other Approaches

- stick with DWT, but use parametric or block bootstrap on each subvector $\mathbf{W}_j$ of coefficients
- for FD processes, $\mathbf{W}_j$ is close to white noise, but the variation from white noise is captured to a very good approximation by an AR(2) process

References: I

- C. Angelini, D. Cava, G. Katul and B. Vidakovic (2005), ‘Resampling Hierarchical Processes in the Wavelet Domain: A Case Study Using Atmospheric Turbulence,’ *Physica D*, **207**, pp. 24-40
- M. Breakspear, M. J. Brammer, E. T. Bullmore, P. Das and L. M. Williams (2004), ‘Spatiotemporal Wavelet Resampling for Functional Neuroimaging Data,’ *Human Brain Mapping*, **23**, pp. 1–25
- M. Breakspear, M. Brammer and P. A. Robinson (2003), ‘Construction of Multivariate Surrogate Sets from Nonlinear Data Using the Wavelet Transform,’ *Physica D*, **182**, pp. 1–22
- E. Bullmore, J. Fadili, V. Maxim, L. Şendur, B. Whitcher, J. Suckling, M. Brammer and M. Breakspear (2004), ‘Wavelets and Functional Magnetic Resonance Imaging of the Human Brain,’ *NeuroImage*, **23**, pp. S234-S249
- E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter and M. Brammer (2001), ‘Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains,’ *Human Brain Mapping*, **12**, pp. 61–78

References: II

- A. C. Davison and D. V. Hinkley (1997), *Bootstrap Methods and their Applications*, Cambridge University Press
- H. Feng, T. R. Willemain and N. Shang (2005), 'Wavelet-Based Bootstrap for Time Series Analysis,' *Communications in Statistics: Simulation and Computation* **34**(2), pp. 393–413
- S. Golia (2002), 'Evaluating the GPH Estimator via Bootstrap Technique,' in *Proceedings in Computational Statistics COMPSTAT2002*, edited by W. Härdle and B. Ronz. Heidelberg: Physica–Verlag, pp. 343–8
- D. B. Percival, S. Sardy and A. C. Davison (2001), 'Wavestrapping Time Series: Adaptive Wavelet-Based Bootstrapping,' in *Nonlinear and Nonstationary Signal Processing*, edited by W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young. Cambridge, England: Cambridge University Press, pp. 442–70
- A. M. Sabatini (1999), 'Wavelet-Based Estimation of $1/f$-Type Signal Parameters: Confidence Intervals Using the Bootstrap,' *IEEE Transactions on Signal Processing*, **47**(12), pp. 3406–9
- B. J Whitcher, 'Wavelet-Based Bootstrapping of Spatial Patterns on a Finite Lattice,' *Computational Statistics & Data Analysis* **50**(9), pp. 2399–421