

Asymptotic Decorrelation of Between-Scale Wavelet Coefficients

Peter F. Craigmile¹ and Donald B. Percival²

¹ Department of Statistics, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus. OH 43210.

² Applied Physics Laboratory, Box 355640, University of Washington, Seattle. WA 98195–5640.

Abstract

In recent years there has been much interest in the analysis of time series using a discrete wavelet transform (DWT) based upon a Daubechies (1992) wavelet filter. Part of this interest has been sparked by the fact that the DWT approximately decorrelates certain stochastic processes, including stationary fractionally differenced (FD) processes with long memory characteristics and certain nonstationary processes such as fractional Brownian motion. In this paper we show that, as the width of the wavelet filter used to form the DWT increases, the covariance between wavelet coefficients associated with different scales decreases to zero for a wide class of stochastic processes. These processes are Gaussian with a spectral density function (SDF) that is the product of the SDF for a (not necessarily stationary) FD process multiplied by any bounded function that can serve as an SDF on its own. We demonstrate that this asymptotic theory provides a reasonable approximation to the between-scale covariance properties of wavelet coefficients based upon filter widths in common use. Our main result is one important piece of an overall strategy for establishing asymptotic results for certain wavelet-based statistics.

1 Introduction

The discrete wavelet transform (DWT) as formulated by Mallat (1989) and Daubechies (1992) is an increasingly popular tool for the statistical analysis of time series (see, e.g., Ogden, 1997; Carmona, Hwang, and Torr sani, 1998; Vidakovic, 1998; Percival and Walden, 2000 and references therein). The DWT maps a time series into a set of wavelet coefficients. Each coefficient is associated with a particular scale, which is a measure of the amount of data that effectively determines the coefficient. Two distinct wavelet coefficients can be either ‘within-scale’ (i.e., both are associated with the same scale) or ‘between-scale’ (i.e., each has a distinct scale).

One reason for the popularity of the DWT is that it acts as a decorrelating (or ‘whitening’) transform for certain stochastic processes; i.e., any two distinct within-scale or between-scale coefficients are approximately uncorrelated. Flandrin (1992) showed that the within-scale wavelet coefficients for fractional Brownian motion (FBM) constitute a stationary process and derived expressions for the covariances of both within-scale and between-scale coefficients. He found that, for the Haar transform and for ordinary Brownian motion, within-scale wavelet coefficients are uncorrelated, while between-scale coefficients are correlated, with the correlation decaying as the separation between the scales increases. Tewfik and Kim (1992) found that wavelet coefficients for FBM are in general correlated, but that the correlation dies down hyperbolically at a rate dictated by the number of vanishing moments associated with the wavelet. Dijkerman and Mazumdar (1994) showed that correlations in the wavelet coefficients for FBM decay exponentially fast between scales and hyperbolically fast within scales. They illustrated that the covariances of distinct wavelet coefficients are small in magnitude. Masry (1993) investigated the statistical properties of the wavelet coefficients of processes with stationary increments (of which FBM is an example). In particular he obtained expressions for the spectral density function (SDF) of within-scale wavelet coefficients and the cross SDF of between-scale coefficients. McCoy and Walden (1996) demonstrated similar results for fractionally differenced (FD) processes, and Vannucci and Corradi (1999) extended the theory of Tewfik and Kim (1992) to these processes (they also introduced a method of calculating the covariances between wavelet coefficients based upon a two dimensional wavelet transform). Jensen (2000) provided the theory for the class of autoregressive fractionally integrated moving average (ARFIMA) processes, of which the FD processes are a specific example. Johnstone and Silverman (1997) and Johnstone (1999) demonstrated the approximate decorrelation property when the Meyer wavelet is used to analyze a process with long range dependence. They used these results to develop wavelet shrinkage techniques for signal estimation in the presence of long range dependence.

In this paper we demonstrate that, for a quite general class of stochastic processes, the covariance of between-scale wavelet coefficients decreases to zero as the width of the wavelet filter associated with the DWT increases (the filter width dictates how much of the time series is used to form each wavelet coefficient: as the width increases, more and more of the series is used to form each coefficient). This result applies even to processes for which the DWT does not yield approximately uncorrelated within-scale coefficients. The theory we present here hence both complements and

generalizes the results given in the literature cited above.

The remainder of this paper is organized as follows. Prior to stating our main result as Theorem 5.1 in §5, we give some background material on processes with stationary backward differences in §2, on fractionally differenced processes in §3 and on wavelet coefficients of stochastic processes in §4. We present a practical demonstration in §6 that this asymptotic result is a good approximation for wavelet filters used in practice. We conclude with a brief discussion in §7 of how our theorem can be used to establish certain consistency results for wavelet-based statistical analysis and how it compares to a well-known decorrelation result in spectral analysis.

2 Processes with Stationary Backward Differences

The decorrelation result we provide in section 5 applies to a particular class of stochastic processes with stationary backward differences. Such processes are introduced in Yaglom (1958) and discussed in detail in Chapter 4 of Yaglom (1987). Notable time series methods for such processes (in the context of long range dependence) can be found in Hurvich and Ray (1995), Beran et al. (1998), and Velasco (1999). Let $\{X_t, t \in \mathbb{Z}\}$ denote a stochastic process, where \mathbb{Z} is the set of all integers. We say that this process has stationary backward differences of integer order $d \geq 0$ if

$$Y_t = \sum_{k=0}^d \binom{d}{k} (-1)^k X_{t-k}$$

is a stationary process with mean μ_Y , autocovariance sequence (ACVS) $\{s_{Y,\tau}, \tau \in \mathbb{Z}\}$ and spectral density function (SDF) $S_Y(\cdot)$ satisfying

$$\int_{-1/2}^{1/2} S_Y(f) e^{i2\pi f\tau} df = s_{Y,\tau}. \quad (1)$$

If $\{X_t\}$ is itself a stationary process with SDF $S_X(\cdot)$, then the theory of linear time-invariant filters (see, e.g., Chapter 5 of Percival and Walden, 1993) dictates that the SDF of $\{Y_t\}$ is

$$S_Y(f) = \mathcal{D}^d(f) S_X(f), \quad (2)$$

where $\mathcal{D}(\cdot)$ is the squared gain function for a first order (i.e., $d = 1$) backward difference filter. By definition, the squared gain function is the squared modulus of the transfer function for the filter.

In turn, the transfer function is the discrete Fourier transform of the sequence defining the filter, which in this case is $a_0 = 1$, $a_1 = -1$ and $a_l = 0$ otherwise. Hence

$$\mathcal{D}(f) = \left| \sum_{l=-\infty}^{\infty} a_l e^{-i2\pi fl} \right|^2 = \left| 1 - e^{-i2\pi f} \right|^2 = 4 \sin^2(\pi f).$$

If $\{X_t\}$ is not a stationary process, we can use Equation (2) to define an SDF for it via the equation

$$S_X(f) \equiv S_Y(f)/\mathcal{D}^d(f) \tag{3}$$

(Yaglom, 1958).

3 Fractionally Differenced Processes

A simple example of a process with stationary backward differences is obtained by letting $S_Y(\cdot)$ in Equation (3) be the SDF for a white noise process with variance σ_ϵ^2 ; i.e., $S_Y(f) = \sigma_\epsilon^2$. In this case, the process $\{X_t\}$ has an SDF is given by

$$S_X(f) = \sigma_\epsilon^2 / \mathcal{D}^d(f) = \sigma_\epsilon^2 / |4 \sin^2(\pi f)|^d.$$

The d th order backward difference for $\{X_t\}$ is a white noise process. A generalization of this example is to replace the nonnegative integer d with any real-valued number δ :

$$S_X(f) = \sigma_\epsilon^2 / |4 \sin^2(\pi f)|^\delta, \quad |f| \leq 1/2. \tag{4}$$

The above can be taken as the definition for a fractionally differenced (FD) process. This process has two parameters, namely, δ and σ_ϵ^2 , which are commonly called the FD parameter and the innovations variance in the literature. This process was originally proposed by Granger and Joyeux (1980) and Hosking (1981) and has become popular in recent years, mainly due to its tractable mathematical properties. When $\delta > 0$, an FD process exhibits ‘long memory’ dependence; when $\delta = 0$, the process reduces to white noise; and, when $\delta < 0$, the process is said to be anti-persistent. An FD process is stationary when $\delta < 1/2$, with an ACVS $\{s_{X,\tau}, \tau \in \mathbb{Z}\}$ that can be readily computed at lag $\tau = 0$ using

$$s_{X,0} = \sigma_\epsilon^2 \frac{\Gamma(1 - 2\delta)}{\Gamma^2(1 - \delta)} \tag{5}$$

and at positive lags using the recursion

$$s_{X,\tau} = s_{X,\tau-1} \frac{\tau + \delta - 1}{\tau - \delta}, \quad \tau = 1, 2, \dots \quad (6)$$

(for negative lags τ , recall that $s_{X,-\tau} = s_{X,\tau}$). When $\delta \geq 1/2$, an FD process has stationary backward differences of order $d = \lfloor \delta + 1/2 \rfloor$ (here $\lfloor x \rfloor$ refers to the greatest integer less than or equal to x).

Further examples of processes with stationary backward differences can be obtained by considering a stochastic process $\{X_t\}$ whose SDF can be expressed as

$$S_X(f) = B(f)/|4 \sin^2(\pi f)|^\delta, \quad (7)$$

where $B(\cdot)$ is a bounded function that is the SDF for some stationary process; i.e., $S_X(\cdot)$ is the product of two SDFs, one for an FD process, and another that is bounded. By letting $B(f) = \sigma_\epsilon^2$ the above reduces to an FD process. Another choice is to let

$$B(f) = \sigma_\epsilon^2 \cdot \frac{|1 - \sum_{k=1}^q \theta_k e^{-i2\pi f k}|^2}{|1 - \sum_{k=1}^p \phi_k e^{-i2\pi f k}|^2},$$

which is bounded if $\{\phi_k\}$ is chosen suitably. This leads to the class of autoregressive, fractionally integrated, moving average (ARFIMA) processes, which by definition possess an SDF given by

$$S_X(f) = \frac{\sigma_\epsilon^2}{|4 \sin^2(\pi f)|^\delta} \cdot \frac{|1 - \sum_{k=1}^q \theta_k e^{-i2\pi f k}|^2}{|1 - \sum_{k=1}^p \phi_k e^{-i2\pi f k}|^2}, \quad |f| \leq 1/2, \quad (8)$$

where $\{\theta_k\}$ and $\{\phi_k\}$ are parameters associated with, respectively, the moving average and autoregressive parts of the process. FD processes can be considered to be a subclass of ARFIMA processes. Beran (1994), Samorodnitsky and Taqqu (1994), and Section 13.2 of Brockwell and Davis (1991) discusses further properties of FD and ARFIMA processes. For a survey of methods of analysis for these processes see Adler et al. (1998), Rangarajan and Ding (2003) and Doukhan et al. (2003).

4 Wavelet Coefficients of Stochastic Processes

Let $\{h_{1,l}, l = 0, \dots, L-1\}$ denote a Daubechies wavelet filter of necessarily even width L (Daubechies (1992)). We presume that $h_{1,0} \neq 0$, $h_{1,L-1} \neq 0$, and $h_{1,l} = 0$ for $l < 0$ or $l \geq L$. By definition, this filter has a squared gain function given by

$$\mathcal{H}_{1,L}(f) \equiv 2 \sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f). \quad (9)$$

Let $g_{1,l} = (-1)^{l+1}h_{1,L-1-l}$ define the associated scaling filter, for which the squared gain function is given by $\mathcal{G}_{1,L}(f) \equiv \mathcal{H}_{1,L}(1/2 - f)$. By Daubechies' construction, we have

$$\mathcal{G}_{1,L}(f) + \mathcal{H}_{1,L}(f) = 2 \text{ for all } f. \quad (10)$$

The wavelet filter is an approximation to an ideal high-pass filter with a passband given by $|f| \in [1/4, 1/2]$, whereas the scaling filter approximates an ideal low-pass filter with passband $|f| \in [0, 1/4]$. We note that there are multiple wavelet filters that share the same squared gain function. These filters have distinct transfer functions, which can be expressed as $H_{1,L}(f) = \mathcal{H}_{1,L}^{1/2}(f) \exp(i\theta_{1,L}(f))$, where $\theta_{1,L}(\cdot)$ denotes the phase function. By requiring the phase function to satisfy certain criteria, Daubechies (1992) defined two widely used families of wavelet filters, namely, the extremal phase filters and the least asymmetric filters. In what follows, we denote the member of width L from these two families as, respectively, the $D(L)$ and $LA(L)$ wavelet filters.

Let $\{X_t, t \in \mathbb{Z}\}$ denote a stochastic process. With $V_{0,t} \equiv X_t$, we can recursively use the equations

$$W_{j,t} \equiv \sum_{l=0}^{L-1} h_{1,l} V_{j-1,2t+1-l} \text{ and } V_{j,t} \equiv \sum_{l=0}^{L-1} g_{1,l} V_{j-1,2t+1-l}, \quad j = 1, 2, \dots,$$

to define the j th level wavelet coefficients $\{W_{j,t}\}$ and scaling coefficients $\{V_{j,t}\}$. The former are associated with changes in averages over a scale of 2^{j-1} and with times spaced 2^j units apart; the latter, with averages over a scale of 2^j having the same spacing in time. With $L_j \equiv (2^j - 1)(L - 1) + 1$, equivalent definitions for these stochastic processes are

$$W_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(t+1)-1-l} \text{ and } V_{j,t} \equiv \sum_{l=0}^{L_j-1} g_{j,l} X_{2^j(t+1)-1-l}, \quad (11)$$

where $\{h_{j,l}, l = 0, \dots, L_j - 1\}$ and $\{g_{j,l}, l = 0, \dots, L_j - 1\}$ are j th level wavelet and scaling filters having transfer functions given by, respectively,

$$H_{j,L}(f) \equiv H_{1,L}(2^{j-1}f) \prod_{k=0}^{j-2} G_{1,L}(2^k f) \text{ and } G_{j,L}(f) \equiv \prod_{k=0}^{j-1} G_{1,L}(2^k f) \quad (12)$$

(see, e.g., Percival and Walden, 2000). The corresponding squared gain functions are given by

$$\mathcal{H}_{j,L}(f) \equiv |H_{j,L}(f)|^2 = \mathcal{H}_{1,L}(2^{j-1}f) \prod_{k=0}^{j-2} \mathcal{G}_{1,L}(2^k f) \text{ and } \mathcal{G}_{j,L}(f) \equiv |G_{j,L}(f)|^2 = \prod_{k=0}^{j-1} \mathcal{G}_{1,L}(2^k f). \quad (13)$$

The following technical lemma is needed in the proof of Theorem 5.1 and in carrying out the calculations reported in §6 (all proofs are in the appendix). The lemma is useful by itself since it

provides an integral expression for calculating the covariance between two wavelet coefficients on possibly different scales, expressed in terms of transfer functions for the wavelet filters at each scale, and the SDF of the $\{X_t\}$ process. The integral can be computed easily in terms of a cross product of the autocovariance for the differenced process, $\{Y_t\}$, with the cumulative sum of the wavelet filters at both wavelet scales. In Equation (14) below, the asterisk denotes complex conjugation.

Lemma 4.1 *Let $\{X_t\}$ be a stochastic process with d th order stationary backward differences $\{Y_t\}$. Let $S_X(\cdot)$ denote the SDF for $\{X_t\}$, and let $\{s_{Y,\tau}\}$ denote the ACVS for $\{Y_t\}$. As defined by Equation (11), let $\{W_{j,t}\}$ and $\{W_{j',t}\}$ be the level j and j' wavelet coefficients for $\{X_t\}$ based upon a wavelet filter $\{h_{1,l}\}$ of width $L \geq 2d$ with transfer function $H_{1,L}(\cdot)$. For all possible levels j and j' and all t and t' , we have*

$$\text{cov}\{W_{j,t}, W_{j',t'}\} = \int_{-1/2}^{1/2} H_{j,L}(f) H_{j',L}^*(f) S_X(f) e^{-i2\pi f(2^j(t+1) - 2^{j'}(t'+1))} df \quad (14)$$

$$= \sum_{l=0}^{L_j-d-1} \sum_{l'=0}^{L_{j'}-d-1} b_{j,l}^{(d)} b_{j',l'}^{(d)} s_{Y,2^j(t+1)-2^{j'}(t'+1)+l-l'}, \quad (15)$$

where $H_{j,L}(\cdot)$ is the transfer function defined in Equation (12) for the j th level wavelet filter $\{h_{j,l}\}$, while $\{b_{j,k}^{(d)}\}$ is the d th order cumulative summation of $\{h_{j,l}\}$; i.e., with $b_{j,l}^{(0)} \equiv h_{j,l}$, we have, for $k = 1, \dots, d$,

$$b_{j,l}^{(k)} = \sum_{m=0}^l b_{j,m}^{(k-1)}, \quad l = 0, \dots, L_j - k - 1.$$

5 Decorrelation of Wavelet Coefficients Between Scales

We now state our main result, which in effect says that the wavelet coefficients on different scales for a quite general class of stochastic processes are asymptotically uncorrelated as the wavelet filter width L increases.

Theorem 5.1 *Let $\{X_t\}$ be a stochastic process possessing an SDF that can be expressed as $S_X(f) = B(f)/|4 \sin^2(\pi f)|^\delta$, where $B(\cdot)$ is a bounded function that is the SDF for some stationary process. As defined by Equation (11), let $\{W_{j,t}\}$ and $\{W_{j',t}\}$ be the level j and j' wavelet coefficients for $\{X_t\}$ based upon a wavelet filter $\{h_{1,l}\}$ of width L with transfer function $H_{1,L}(\cdot)$. Then for $j \neq j'$, $\text{cov}\{W_{j,t}, W_{j',t'}\} \rightarrow 0$ at rate $L^{-1/4}$ as $L \rightarrow \infty$.*

This (possibly conservative) rate of decorrelation is indicated by the proof of the theorem. The following heuristic argument says that the above theorem is plausible. Let

$$\mathcal{H}_I(f) \equiv \begin{cases} 0, & |f| < 1/4; \\ 1, & f = 1/4; \text{ and} \\ 2, & 1/4 < |f| \leq 1/2, \end{cases}$$

and define $\mathcal{H}_I(\cdot)$ outside of the interval $[-1/2, 1/2]$ so that it is a periodic function with unit period corresponding to the squared gain function for an ideal high-pass filter. Lai (1995) shows that $\mathcal{H}_{1,L}(f) \rightarrow \mathcal{H}_I(f)$ as $L \rightarrow \infty$. By making use of Equations (10) and (13), we can argue that $\mathcal{H}_{j,L}(\cdot)$ converges to the squared gain function for a ideal band-pass filter with pass-band given by $|f| \in [2^{-(j+1)}, 2^{-j}]$. For $j = 1, 2, \dots$, these ideal squared gain functions disjointly partition the frequency interval $[-1/2, 1/2]$. The spectral representation theorem for zero mean stationary processes states that

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft} dZ(f),$$

where $\{Z(f)\}$ is a process with orthogonal increments (Cramér, 1942). For large L , we thus have

$$W_{j,t} \approx \int_{|f| \in [2^{-(j+1)}, 2^{-j}]} H_{j,L}(f) e^{i2\pi ft} dZ(f).$$

The pass-bands for wavelet coefficients do not intersect, and hence the fact that $\{Z(f)\}$ has uncorrelated increments implies that $W_{j,t}$ and $W_{j',t'}$ are uncorrelated when $j \neq j'$.

6 Examples of the Decorrelation Property

Here we demonstrate that the asymptotic result stated in Theorem 5.1 is in fact a good approximation for wavelet filters used in practice. Figure 1 shows plots of the maximum absolute correlation between $W_{j,t}$ and $W_{j',t'}$ for a fixed pair (j, j') of levels satisfying $1 \leq j < j' \leq 4$ as t and t' range over all possible integers. The process from which the wavelet coefficients are formed is an FD process with $\delta = 0.4$, and the wavelet filters are either the $D(L)$ (pluses) or $LA(L)$ filters (circles) of widths $L = 2, 4, \dots, 20$ (the case $L = 2$ for both families corresponds to the Haar wavelet filter, and the $LA(4)$ and $LA(6)$ filters are the same as the $D(4)$ and $D(6)$ filters). This particular FD process is cited in Percival and Walden (2000) as an example of one for which the DWT acts as a decent within-scale prewhitening transform. The plots in the figure indicate that the maximum

absolute between-scale correlation decreases as L increases in keeping with Theorem 5.1. Whereas the maximum correlation for the Haar wavelet ($L = 2$) can be as high as 0.2 ($j = 2, j' = 4$), this correlation decreases to at most 0.03 ($j = 3, j' = 4$) when we use a filter of width $L = 20$. This example also illustrates that, for larger $|j - j'|$, we get faster decorrelation as $L \rightarrow \infty$.

Figure 2 shows a second example, now for an autoregressive process of unit order given by $X_t = -0.9X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is a zero mean Gaussian white noise process. This process is cited in Percival and Walden (2000) as an example of one for which the wavelet coefficients within unit scale are not approximately white noise. Again we can see the between-scale correlations decrease as we increase L , but this decrease need not be monotonic ($j = 2$ and 3). Thus, even though the DWT might not be an adequate within-scale decorrelator for a particular process, we can still achieve decent between-scale decorrelation by increasing L , again in keeping with Theorem 5.1.

We note one pathological process (not typical of time series seen in practical applications) for which the covariance between scales decreases to zero with increasing L , but the correlation between certain scales does not. Consider a Gaussian white noise process that is bandlimited to the frequency range $(-1/8, 1/8)$; i.e., the SDF is identically zero for $|f| \in [1/8, 1/2]$. The conditions for Theorem 5.1 hold, so the covariance between wavelet coefficients on levels 1 and 2 does decrease to zero; however, a plot similar to those in Figures 1 and 2 indicates correlations close to unity for all L considered. The reason for the high correlation is that, because there is no power in the nominal pass-bands associated with levels 1 and 2, both sets of wavelet coefficients essentially depend on leakage from the same frequency range $(-1/8, 1/8)$ and hence are coupled together.

7 Implications for Time Series Analysis

In this paper we have established that, for a quite general class of stochastic processes, the covariance of between-scale wavelet coefficients decreases as we increase the width L of the wavelet filter. It is not possible to establish a similar general result for show within-scale wavelet coefficients. In fact increasing L can actually lead to an increase in the covariance of these coefficients, a fact that is documented in Percival and Walden (2000). The statistical implication and utility of these facts is that, in order to study the exact asymptotic properties of certain wavelet-based statistics, we can let

L get large in order to obtain approximately uncorrelated between-scale coefficients, but we are then faced with the problem of modeling the within-scale covariance structure (this same difficulty arises in the context of wavelet thresholding, where the use of block thresholding techniques also requires an understanding of the within-scale structure; see, e.g., Cai and Silverman, 2001). In previous work (Craigmile et al, 2000; Craigmile et al, 2004), we have found that, for processes for which the DWT does a reasonable job of decorrelating within scales (e.g., an FD process), we can model the remaining within-scale covariance quite well using an autoregressive (AR) process of order $p = 1$ or 2 . Now a wide class of stationary processes can be approximated arbitrarily well using an $\text{AR}(p)$ process of high enough order. This suggests that an overall approach to a useful asymptotic theory for statistics based on wavelet coefficients is to let $L \rightarrow \infty$ and $p \rightarrow \infty$. This decorrelation result gives some asymptotic justification for wavelet-based cross validation (Nason, 1996), wavelet-based least square estimation methods (Abry et al, 1993; Abry et al, 1995; Veitch and Abry, 1999; Bardet et al, 2000), and wavelet-based bootstrapping methods (Sabatini, 1999; Percival et al, 2000). This is because, for large enough L we can consider the different wavelet scales to be approximately uncorrelated, and in the Gaussian case we can resample from each wavelet scale (almost) independently of other scales. The result also gives guidance on designing wavelet-based methods for the simulation of Gaussian processes (e.g., Percival and Walden, 2000, Section 9.4).

It is of interest to contrast our basic result to a similar one arising in spectral analysis. Suppose we observe a time series that is a realization of $\{X_t : t \in \mathbb{Z}\}$ belonging to the class of processes given in Theorem 5.1; i.e., $\{X_t\}$ is a stochastic process with an SDF that can be expressed as $S_X(f) = B(f)/|4 \sin^2(\pi f)|^\delta$, where $B(\cdot)$ is a bounded function that is the SDF for some stationary process. The most popular estimator of the SDF is the periodogram, which is based upon the squared magnitude of the coefficients of the discrete Fourier transform (DFT):

$$\widehat{S}^{(p)}(f_k) = \frac{1}{N} \left| \sum_{t=0}^{N-1} X_t e^{-2\pi f_k t} \right|^2,$$

where $f_k = k/N$ denotes the k th Fourier frequency, $k = 0, \dots, \lfloor N/2 \rfloor$. Consider two sequences of Fourier frequencies f_j and f_k such that $f_j \rightarrow f$ and $f_k \rightarrow f'$ as $N \rightarrow \infty$, where f and f' are fixed frequencies such that $0 < f < f' < 1/2$. Under these conditions, results in Beran (1994) and Yajima (1989) indicate that $\widehat{S}^{(p)}(f_j)$ and $\widehat{S}^{(p)}(f_k)$ are asymptotically uncorrelated as $N \rightarrow \infty$ when $\delta \in [-1/2, 1/2)$. This result is similar to between-scale decorrelation of wavelet coefficients, but there

are some key differences. First, asymptotic between-scale decorrelation is in terms of filter width L rather than sample size N and is a reasonable approximation for filter widths used in practice. By contrast, even for very large N , the periodogram can suffer from severe bias, which invalidates basic properties suggested by the asymptotic theory. Second, the theory for wavelet coefficients is applicable when $\delta \geq 1/2$, which is not the case for the theory stated above for spectral analysis. In general, the differencing operations that are embedded in the Daubechies wavelet filters make it an easy matter for the DWT to handle nonstationary processes with stationary increments (i.e., $\delta \geq 1/2$) in a tractable manner. This stems from the fact that, if L is properly matched to the order of differencing required to reduce the process to stationarity, the resulting wavelet coefficients can be described as a portion of a zero mean stationary process. The situation is quite different for DFT coefficients. Because the DFT does not have embedded differencing operations and because all DFT coefficients involve the entire time series and can thus be adversely influenced by mismatches between the beginning and end of the series, the DFT coefficients cannot be so easily described.

We close with a practical note. We have cast our theoretical development in terms of stochastic processes defined over all integers. In practical applications, we must deal with a time series that can be regarded as one realization of a finite portion X_0, \dots, X_{N-1} of such a process. In the most common implementation of the algorithm, the DWT of a time series involves a number of coefficients that are circularly filtered, i.e., that combine together values from the beginning and end of the time series. Other implementations of the DWT involve the use of reflection boundary conditions or boundary wavelets. Regardless of the implementation chosen, these ‘boundary’ wavelet coefficients can have markedly different statistical properties from those extracted from the corresponding theoretical stochastic process. As L increases, the number of boundary coefficients on each scale increases. Use of these boundary coefficients can result in biases in certain wavelet-based statistics. Clearly, before increasing L , there is a need to carefully study how the benefit from a decrease in the between-scale covariances is impacted by having fewer nonboundary coefficients to work with (this is a common problem in most statistical applications when one filters data).

Acknowledgements

The authors gratefully acknowledges support for this research from an STTR grant from AFOSR (Insightful, Inc., and the University of Washington), NSF grant No. 0222115 (University of Washington), and an EPA grant (National Research Center for Statistics and the Environment). We would also like to thank Fadoua Balabda for discussions concerning the proofs of Proposition 8.2 and Lemma 8.3, and for helpful comments by Prof. Noel Cressie. We also extend our thanks to the Associate Editor and three referees, who gave suggestions that improved this paper.

8 Proofs

Proposition 8.1 is needed in the proof of Lemma 4.1.

Proposition 8.1 *Let $\{h_{j,l}\}$ be the j th level wavelet filter of width $L_j \equiv (2^j - 1)(L - 1) + 1$ formed using the Daubechies unit level wavelet filter $\{h_{1,l}\}$ of width L . For any $0 \leq d \leq L/2$, we can write*

$$h_{j,l} = \sum_{k=0}^d \binom{d}{k} (-1)^k b_{j,l-k}^{(d)}, \quad (16)$$

where the filter $\{b_{j,l}^{(d)}\}$ is defined to be the d th successive cumulative summation of $\{h_{j,l}\}$ and is of width $L_j - d$.

Proof of Proposition 8.1 Equation (16) holds trivially when $d = 0$ because $b_{j,l}^{(0)} = h_{j,l}$ in this case. Suppose now that it holds for some $0 \leq d < L/2$. We claim that it must hold for $d + 1$ also. To establish the claim, we must show that, given (16), we have

$$h_{j,l} = \sum_{k=0}^{d+1} \binom{d+1}{k} (-1)^k b_{j,l-k}^{(d+1)} = (1 - B)^{d+1} b_{j,l}^{(d+1)},$$

where B is the backward shift operator (i.e., $B b_{j,l}^{(d+1)} = b_{j,l-1}^{(d+1)}$ and $B^k b_{j,l}^{(d+1)} = b_{j,l-k}^{(d+1)}$ for any nonnegative integer k). Starting with the definition of $b_{j,l}^{(1)}$ and using Equation (16), we have

$$\begin{aligned} b_{j,l}^{(1)} &\equiv \sum_{m=-\infty}^l h_{j,m} = \sum_{k=0}^d \binom{d}{k} (-1)^k \sum_{m=-\infty}^l b_{j,m-k}^{(d)} = \sum_{k=0}^d \binom{d}{k} (-1)^k \sum_{m=-\infty}^{l-k} b_{j,m}^{(d)} \\ &= (1 - B)^d \sum_{m=-\infty}^l b_{j,m}^{(d)} = (1 - B)^d b_{j,l}^{(d+1)} \end{aligned}$$

because $\{b_{j,l}^{(d+1)}\}$ is the cumulative sum of $\{b_{j,l}^{(d)}\}$. Hence

$$h_{j,l} = \sum_{m=-\infty}^l h_{j,m} - \sum_{m=-\infty}^{l-1} h_{j,m} = b_{j,l}^{(1)} - b_{j,l-1}^{(1)} = (1-B)b_{j,l}^{(1)} = (1-B)^{d+1}b_{j,l}^{(d+1)},$$

as required.

To complete the proof of the proposition, we need to show that $\{b_{j,l}^{(d)}\}$ has width $L_j - d$, for which we will also use an argument by induction. Since $b_{j,l}^{(0)} = h_{j,l}$, the result holds trivially for $d = 0$. Assume now that $\{b_{j,l}^{(d)}\}$ has width $L_j - d$ for some $0 \leq d < L/2$. Since $\{b_{j,l}^{(d+1)}\}$ is the cumulative sum of $\{b_{j,l}^{(d)}\}$, it follows that the former has width $L_j - d - 1$ if we can show that

$$b_{j,L_j-d-1}^{(d+1)} = \sum_{m=0}^{L_j-d-1} b_{j,m}^{(d)} = B_j^{(d)}(0) = 0,$$

where $B_j^{(d)}(\cdot)$ is the transfer function for $\{b_{j,l}^{(d)}\}$. From the solution to Exercise [304], Percival and Walden (2000), the transfer function for $\{h_{j,l}\}$ can be written as $H_{j,L}(f) = D^{L/2}(f)A_j(f)$, where $D(f) = 1 - e^{-i2\pi f}$ is the transfer function for the first order difference filter, while $A_j(\cdot)$ is a continuous periodic function. Equation (16) says that we also have $H_{j,L}(f) = D^d(f)B_j^{(d)}(f)$. Since $D(f) \neq 0$ except at $f = 0$, we must have $B_j^{(d)}(f) = D^{L/2-d}(f)A_j(f)$ for all $f \neq 0$. By continuity,

$$B_j^{(d)}(0) = \lim_{f \rightarrow 0} B_j^{(d)}(f) = \lim_{f \rightarrow 0} D^{L/2-d}(f) \cdot \lim_{f \rightarrow 0} A_j(f) = D^{L/2-d}(0)A_j(0) = 0$$

because $L/2 > d$, and hence $D^{L/2-d}(0) = 0$. \square

Proof of Lemma 4.1 The case $d = 0$ follows from Exercise [348a] of Percival and Walden (2000), so we assume $d \geq 1$. For convenience, define $\binom{d}{k} = 0$ when $k < 0$ or $k > d$ so that summations over k below can be taken to range over all integers. It follows from the definition for $W_{j,t}$ in Equation (11) and from Proposition 8.1 that

$$\begin{aligned} W_{j,t} &= \sum_l \left[\sum_k \binom{d}{k} (-1)^k b_{j,l-k}^{(d)} \right] X_{2^j(t+1)-1-l} \\ &= \sum_l \left[\sum_k \binom{d}{l-k} (-1)^{l-k} b_{j,k}^{(d)} \right] X_{2^j(t+1)-1-l} \\ &= \sum_k \left[\sum_l \binom{d}{l-k} (-1)^{l-k} X_{2^j(t+1)-1-l} \right] b_{j,k}^{(d)} \\ &= \sum_k \left[\sum_l \binom{d}{l} (-1)^l X_{2^j(t+1)-1-l-k} \right] b_{j,k}^{(d)} = \sum_k b_{j,k}^{(d)} Y_{2^j(t+1)-1-k}, \end{aligned}$$

from which Equation (15) follows using the basic result

$$\text{cov} \left\{ \sum_j c_j U_j, \sum_k d_k V_k \right\} = \sum_j \sum_k c_j d_k \text{cov} \{U_j, V_k\}.$$

As before, let $B_j^{(d)}(\cdot)$ and $D(\cdot)$ denote the transfer functions for, respectively, $\{b_{j,l}^{(d)}\}$ and the first order backward difference filter, and note that $\mathcal{D}(f) = D(f)D^*(f)$. Starting from Equation (15) and making use of Equations (1) and (3), we have

$$\begin{aligned} \text{cov} \{W_{j,t}, W_{j',t'}\} &= \sum_{l=0}^{L_j-d-1} \sum_{l'=0}^{L_{j'}-d-1} b_{j,l}^{(d)} b_{j',l'}^{(d)} \int_{-1/2}^{1/2} S_Y(f) e^{i2\pi f(2^j(t+1)-2^{j'}(t'+1)+l'-l)} df \\ &= \int_{-1/2}^{1/2} \left(\sum_{l=0}^{L_j-d-1} b_{j,l}^{(d)} e^{-i2\pi fl} \right) \left(\sum_{l'=0}^{L_{j'}-d-1} b_{j',l'}^{(d)} e^{-i2\pi fl'} \right)^* \\ &\quad \times S_X(f) e^{i2\pi f(2^j(t+1)-2^{j'}(t'+1))} df \\ &= \int_{-1/2}^{1/2} B_j^{(d)}(f) [B_{j'}^{(d)}(f)]^* S_Y(f) e^{i2\pi f(2^{j'}(t'+1)-2^j(t+1))} df \\ &= \int_{-1/2}^{1/2} D^d(f) B_j^{(d)}(f) [D^d(f) B_{j'}^{(d)}(f)]^* \frac{S_Y(f)}{\mathcal{D}^d(f)} e^{i2\pi f(2^{j'}(t'+1)-2^j(t+1))} df \\ &= \int_{-1/2}^{1/2} H_{j,L}(f) H_{j',L}^*(f) S_X(f) e^{i2\pi f(2^{j'}(t'+1)-2^j(t+1))} df, \end{aligned}$$

which is Equation (14). \square

In preparation for our proof of Theorem 5.1, we state and prove some preliminary results.

Proposition 8.2 *For $f \in [0, 1/2]$, the squared gain function for the scaling filter $\{g_{1,l}\}$ can be expressed as*

$$\mathcal{G}_{1,L}(f) = \frac{2 \int_{2\pi f}^{\pi} \sin^{L-1}(y) dy}{B(L/2, 1/2)},$$

where $B(\cdot, \cdot)$ is the beta function; i.e.,

$$B(x, y) = 2 \int_0^{\pi/2} \sin^{2x-1}(t) \cos^{2y-1}(t) dt.$$

Proof of Proposition 8.2 It follows directly from Equation (5.70), p. 168, Strang and Nguyen (1996), that the Meyer form for $\mathcal{G}_{1,L}(\cdot)$ yields

$$\mathcal{G}_{1,L}(f) = 2 - \frac{2 \int_0^{2\pi f} \sin^{L-1}(y) dy}{\int_0^{\pi} \sin^{L-1}(y) dy}.$$

Thus

$$\mathcal{G}_{1,L}(f) = 2 \left(\frac{\int_0^\pi \sin^{L-1}(y) dy - \int_0^{2\pi f} \sin^{L-1}(y) dy}{\int_0^\pi \sin^{L-1}(y) dy} \right) = \frac{\int_{2\pi f}^\pi \sin^{L-1}(y) dy}{\int_0^{\pi/2} \sin^{L-1}(y) dy} = \frac{2 \int_{2\pi f}^\pi \sin^{L-1}(y) dy}{B(L/2, 1/2)},$$

where, in the next to last step, we have made use of the fact that $\sin^{L-1}(y)$ is symmetric about $y = \pi/2$. \square

Lemma 8.3 $\pi/2 - \arcsin(\sqrt{z}) \leq (\pi/2)(1-z)^{1/2}$ for $z \in [0, 1]$.

Proof of Lemma 8.3 Let $f(z) = \pi/2 - \arcsin(\sqrt{z}) - (\pi/2)(1-z)^{1/2}$. Then

$$f'(z) = -\frac{1}{2\sqrt{z(1-z^2)}} + \frac{\pi}{4\sqrt{1-z}} = \frac{\pi\sqrt{z(1+z)} - 2}{4\sqrt{z(1-z)(1+z)}}.$$

We now solve $f'(z) = 0$ with respect to $z \in [0, 1]$. Since the denominator is nonnegative and finite, we need only examine the numerator $f'_{(n)}(z) \equiv \pi\sqrt{z(1+z)} - 2$. Algebraic manipulations show that the condition $f'_{(n)}(z) = 0$ is equivalent to $z^2 + z - (2/\pi)^2 = 0$. The roots of $f'_{(n)}(z)$ are given by

$$z = -\frac{1}{2} \pm \frac{\sqrt{1 + 16/\pi^2}}{2}.$$

Thus

$$f'_{(n)}(z) = \left(z - \frac{-1 + \sqrt{1 + 16/\pi^2}}{2} \right) \left(z - \frac{-1 - \sqrt{1 + 16/\pi^2}}{2} \right).$$

For $z \in [0, 1]$, $f'_{(n)}(z)$ has a turning point at $z = (-1 + \sqrt{1 + 16/\pi^2})/2 \doteq 0.309$. We have $f'(z)$ is negative below the turning point, and positive above it, implying that $f(z)$ is decreasing below the turning point, and increasing above it. Now $f(0) = \pi/2 - 0 - \pi/2 = 0$ and $f(1) = \pi/2 - \pi/2 - 0 = 0$, yielding the desired result. \square

Theorem 8.4 $\int_0^{1/2} \mathcal{H}_{1,L}(f) \mathcal{G}_{1,L}(f) df \rightarrow 0$ as $L \rightarrow \infty$.

Proof of Theorem 8.4 We have

$$\begin{aligned} \int_0^{1/2} \mathcal{H}_{1,L}(f) \mathcal{G}_{1,L}(f) df &= \int_0^{1/2} \mathcal{G}_{1,L}(1/2 - f) \mathcal{G}_{1,L}(f) df = 2 \int_0^{1/4} \mathcal{G}_{1,L}(1/2 - f) \mathcal{G}_{1,L}(f) df \\ &\leq 4 \int_0^{1/4} \mathcal{G}_{1,L}(1/2 - f) df \end{aligned}$$

by symmetry of $\mathcal{G}_{1,L}(f)$ about $f = 1/4$ and by using the fact (from Equation (10)) that $\mathcal{G}_{1,L}(f) \leq 2$ for all f . To prove the result we need only show that the final integral above goes to zero as $L \rightarrow \infty$.

From proposition 8.2

$$\int_0^{1/4} \mathcal{G}_{1,L}(1/2 - f) df = \frac{2}{B(L/2, 1/2)} \int_0^{1/4} \int_{2\pi(1/2-f)}^{\pi} \sin^{L-1}(y) dy df.$$

Now consider the numerator. Using the change of variable $u = 2\pi f$, we have

$$\int_0^{1/4} \int_{2\pi(1/2-f)}^{\pi} \sin^{L-1}(y) dy df = \frac{1}{2\pi} \int_0^{\pi/2} \int_{\pi-u}^{\pi} \sin^{L-1}(y) dy du = \frac{1}{2\pi} \int_0^{\pi/2} \int_0^u \sin^{L-1}(y) dy du$$

by the symmetry of $\sin^{L-1}(y)$ about $y = \pi/2$. We now make the change of variable $z = \sin^2(y)$.

Then $dz = 2 \sin(y) \cos(y) dy = 2\sqrt{z(1-z)} dy$ and the above expression is equal to

$$\begin{aligned} \frac{1}{2\pi} \int_0^{\pi/2} \int_0^{\sin^2(u)} \frac{z^{(L-1)/2}}{2\sqrt{z(1-z)}} dz du &= \frac{1}{4\pi} \int_0^{\pi/2} \int_0^{\sin^2(u)} z^{L/2-1} (1-z)^{-1/2} dz du \\ &= \frac{1}{4\pi} \int_0^{\pi/2} \int_0^1 z^{L/2-1} (1-z)^{-1/2} I_{[z < \sin^2(u)]} dz du \\ &= \frac{1}{4\pi} \int_0^{\pi/2} \int_0^1 z^{L/2-1} (1-z)^{-1/2} I_{[\arcsin(\sqrt{z}) < u]} dz du \\ &= \frac{1}{4\pi} \int_0^1 z^{L/2-1} (1-z)^{-1/2} \int_{\arcsin(\sqrt{z})}^{\pi/2} du dz \\ &= \frac{1}{4\pi} \int_0^1 z^{L/2-1} (1-z)^{-1/2} \left(\frac{\pi}{2} - \arcsin(\sqrt{z}) \right) dz \\ &\leq \frac{1}{4\pi} \int_0^1 z^{L/2-1} (1-z)^{-1/2} \left(\frac{\pi}{2} \right) (1-z)^{1/2} dz \\ &= \frac{1}{8} \int_0^1 z^{L/2-1} dz = (4L)^{-1}, \end{aligned}$$

where, in the next to last line, we have made use of Lemma 8.3. Thus

$$\int_0^{1/4} \mathcal{G}_{1,L}(1/2 - f) df \leq \frac{1}{2L \cdot B(L/2, 1/2)} = \frac{\Gamma(L/2 + 1/2)}{2L \cdot \Gamma(L/2)\Gamma(1/2)} = \frac{\Gamma(L/2 + 1/2)}{4\Gamma(L/2 + 1)\sqrt{\pi}}.$$

Using Stirling's formula $\Gamma(x+1) \sim \sqrt{2\pi} x^{x+1/2} e^{-x}$, we have

$$\Gamma(L/2 - 1/2 + 1) \sim \sqrt{2\pi} [(L-1)/2]^{L/2} e^{-L/2+1/2} \quad \text{and} \quad \Gamma(L/2 + 1) \sim \sqrt{2\pi} (L/2)^{L/2+1/2} e^{-L/2},$$

from which it follows that

$$\frac{\Gamma(L/2 + 1/2)}{4\Gamma(L/2 + 1)\sqrt{\pi}} \sim \frac{[(L-1)/2]^{L/2}}{(L/2)^{(L+1)/2} 4e^{1/2}\sqrt{\pi}} = \frac{(1-1/L)^{L/2}}{(L/2)^{1/2} 4e^{1/2}\sqrt{\pi}}.$$

Now

$$\log([1 - 1/L]^{L/2}) = \frac{L}{2} \log(1 - 1/L) \sim \frac{L}{2} \cdot \frac{1}{L} = \frac{1}{2},$$

so $(1 - 1/L)^{L/2} \rightarrow \exp(1/2)$ as $L \rightarrow \infty$. Hence

$$\frac{(1 - 1/L)^{L/2}}{(L/2)^{1/2} 4e^{1/2} \sqrt{\pi}} = O(1/L^{1/2}),$$

which goes to zero as $L \rightarrow \infty$. \square

Corollary 8.5 *Suppose that $j \geq 1$. As $L \rightarrow \infty$*

$$\int_0^{1/2} \mathcal{H}_{1,L}(2^{j-1}f) \mathcal{G}_{1,L}(2^{j-1}f) df \rightarrow 0.$$

Proof of Corollary 8.5 Let $u = 2^{j-1}f$. Then

$$2^{-j+1} \int_0^{2^{j-2}} \mathcal{H}_{1,L}(u) \mathcal{G}_{1,L}(u) du = 2^{-j+1} \sum_{j=0}^{2^{j-1}-1} \int_0^{1/2} \mathcal{H}_{1,L}(u) \mathcal{G}_{1,L}(u) du.$$

by the periodicity and evenness of $\mathcal{H}(\cdot)$ and $\mathcal{G}(\cdot)$. The result follows by Theorem 8.4. \square

Proof of Theorem 5.1 Without loss of generality, assume that $j < j'$. Using Equation (14) and the fact that $|e^{ix}| = 1$ for all x , it follows that

$$\begin{aligned} |\text{cov}\{W_{j,t}, W_{j',t'}\}| &\leq \int_{-1/2}^{1/2} |H_{j,L}(f) H_{j',L}^*(f)| S_X(f) df \\ &= 2 \int_0^{1/2} [\mathcal{H}_{j,L}(f) \mathcal{H}_{j',L}(f) S_X^2(f)]^{1/2} df. \end{aligned}$$

It thus suffices to show that the second integral goes to zero as $L \rightarrow \infty$. Since Equation (10) implies that $0 \leq \mathcal{G}_{1,L}(f) \leq 2$ for all f , it follows from Equation (13) that

$$\begin{aligned} \mathcal{H}_{j,L}(f) \mathcal{H}_{j',L}(f) S_X^2(f) &= \left[\mathcal{H}_{1,L}(2^{j-1}f) \prod_{k=0}^{j-2} \mathcal{G}_{1,L}(2^k f) \right] \left[\mathcal{H}_{1,L}(2^{j'-1}f) \prod_{k=0}^{j'-2} \mathcal{G}_{1,L}(2^k f) \right] S_X^2(f) \\ &\leq 2^{j+j'-3} [\mathcal{H}_{1,L}(2^{j-1}f) \mathcal{G}_{1,L}(2^{j-1}f)] [\mathcal{H}_{1,L}(2^{j'-1}f) S_X^2(f)], \end{aligned}$$

where, in the first line above, we interpret the product within the first set of brackets to be unity when $j = 1$. Using the above and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} |\text{cov}\{W_{j,t}, W_{j',t'}\}|^2 &\leq 2^{j+j'-2} \left(\int_0^{1/2} [\mathcal{H}_{1,L}^{1/2}(2^{j-1}f) \mathcal{G}_{1,L}^{1/2}(2^{j-1}f)] \cdot [\mathcal{H}_{1,L}^{1/2}(2^{j'-1}f) S_X(f)] df \right)^2 \\ &\leq 2^{j+j'-2} \left(\int_0^{1/2} \mathcal{H}_{1,L}(2^{j-1}f) \mathcal{G}_{1,L}(2^{j-1}f) df \right) \left(\int_0^{1/2} \mathcal{H}_{1,L}(2^{j'-1}f) S_X^2(f) df \right). \end{aligned}$$

Corollary 8.5 indicates that the first integral in the last line goes to zero as $L \rightarrow \infty$, so the desired results follows if we can show that

$$\int_0^{1/2} \mathcal{H}_{1,L}(2^{j'-1}f) S_X^2(f) df = \int_0^{1/2} \frac{\mathcal{H}_{1,L}(2^{j'-1}f) B^2(f)}{|4 \sin^2(\pi f)|^{2\delta}} df \leq C \quad (17)$$

where C is a finite constant independent of L when L is large enough. Since $B(\cdot)$ is nonnegative and bounded above by, say, C_0 , we have

$$\int_0^{1/2} \mathcal{H}_{1,L}(2^{j'-1}f) S_X^2(f) df \leq C_0^2 \int_0^{1/2} \frac{\mathcal{H}_{1,L}(2^{j'-1}f)}{|4 \sin^2(\pi f)|^{2\delta}} df.$$

If $\delta \leq 0$, the right-hand side is bounded by $2^{-4\delta+1}C_0^2$, so Equation (17) holds in this case. Suppose $\delta > 0$. The change of variable $u = 2^{j'-1}f$ yields

$$\int_0^{1/2} \frac{\mathcal{H}_{1,L}(2^{j'-1}f)}{|4 \sin^2(\pi f)|^{2\delta}} df = 2^{-j'+1} \int_0^{2^{j'-2}} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(2^{-j'+1}\pi u)|^{2\delta}} du.$$

Now $1/|4 \sin^2(2^{-j'+1}\pi u)|^{2\delta}$ is a decreasing function over the interval $[0, 1/2]$, so its maximum value over $u \in [1/2, 2^{j'-2}]$ occurs at $u = 1/2$ and is equal to, say, C_1 . Hence

$$\int_0^{1/2} \mathcal{H}_{1,L}(2^{j'-1}f) S_X^2(f) df \leq 2^{-j'+1} C_0^2 \left(2C_1 + \int_0^{1/2} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(2^{-j'+1}\pi u)|^{2\delta}} du \right),$$

so the desired result follows if

$$\int_0^{1/2} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(2^{-j'+1}\pi u)|^{2\delta}} du \leq C_2$$

for some finite constant C_2 independent of L when L is large enough. Making use of $\sin(2x) = 2 \sin(x) \cos(x)$ yields

$$\sin(2^{-j'+1}\pi u) = \frac{\sin(2^{-j'+2}\pi u)}{2 \cos(2^{-j'+1}\pi u)} \text{ and hence } \sin(2^{-j'+1}\pi u) = \frac{\sin(\pi u)}{2^{j'-1} \prod_{k=1}^{j'-1} \cos(2^{-k}\pi u)}.$$

For $u \in [0, 1/2]$, the product in the denominator is positive and bounded above by $2^{j'-1}$ and hence

$$\int_0^{1/2} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(2^{-j'+1}\pi u)|^{2\delta}} du \leq 2^{4\delta(j'-1)} \int_0^{1/2} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(\pi u)|^{2\delta}} du \equiv \nu_L^2,$$

where ν_L^2 is the unit level wavelet variance for an FD process with fractional difference parameter 2δ and is finite as long as $L \geq \max\{2[2\delta + 1/2], 2\} \equiv L_0$ (Percival and Walden, 2000). Thus, for all $L \geq L_0$,

$$\int_0^{1/2} \frac{\mathcal{H}_{1,L}(u)}{|4 \sin^2(\pi u)|^{2\delta}} du \leq \int_0^{1/6} \frac{\mathcal{H}_{1,L_0}(u)}{|4 \sin^2(\pi u)|^{2\delta}} du + \frac{2}{3} \equiv \frac{C_2}{2^{4\delta(j'-1)}} < \infty,$$

where we have made use of the facts that (i) $\mathcal{H}_{1,L}(f)$ monotonically decreases to zero as L increases for all $f \in (0, 1/6]$ (Lai, 1995); (ii) $\mathcal{H}_{1,L}(u) \leq 2$; (iii) the maximum value of $1/|4 \sin^2(\pi u)|^{2\delta}$ over $u \in [1/6, 1/2]$ occurs at $u = 1/6$; and (iv) $\sin(\pi/6) = 1/2$. \square

References

- Abry, P., P. Gonçalvès, and P. Flandrin (1993). Wavelet-based spectral analysis of $1/f$ processes. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Volume 3, pp. 237–240. Minneapolis, MN, USA.
- Abry, P., P. Gonçalvès, and P. Flandrin (1995). *Wavelets, Spectrum Analysis and $1/f$ processes*, Volume 103 of *Lecture Notes in Statistics*, pp. 15–29. New York: Springer-Verlag.
- Adler, R. J., R. E. Feldman, and M. S. Taqqu (1998). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhuser Verlag.
- Bardet, J., G. Lang, E. Moulines, and P. Soulier (2000). Wavelet estimator of long range dependence processes. *Statistical Inference for Stochastic Processes* 3(1-2), 85–99.
- Beran, J. (1994). *Statistics for Long Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall.
- Beran, J., R. J. Bhansali, and D. Ocker (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika* 85, 921–934.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series. Theory and Methods (Second Edition)*. New York: Springer-Verlag.
- Cai, T. T. and B. W. Silverman (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya, Series B* 63(2), 127–148.
- Carmona, R., W.-L. Hwang, and B. Torrèsani (1998). *Practical Time-Frequency Analysis*. San Diego: Academic Press.
- Craigmile, P. F., D. B. Percival, and P. Guttorp (2000). The impact of wavelet coefficient correlations on fractionally differenced process estimation. In *European Congress of Mathematics: Barcelona, July 10–14, 2000, Volume II*, Progress in Mathematics (Volume 202), Basel, pp. 591–599. Birkhäuser Verlag.
- Craigmile, P. F., D. B. Percival, and P. Guttorp (2004). Wavelet-based parameter estimation for trend contaminated fractionally differenced processes. Technical Report 77, National Research Center for Statistics and the Environment, University of Washington.

- Cramér, H. (1942). On harmonic analysis in certain functional spaces. *Archiv för Matematik, Astonomi och Fysik 28B*, 1–7.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. Philadelphia: SIAM.
- Dijkerman, R. and R. Mazumdar (1994). On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory 40*, 1609–1612.
- Doukhan, P., G. Oppenheim, and M. S. Taqqu (Eds.) (2003). *Theory and Applications of Long-Range Dependence*. Boston: Birkhauser.
- Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory 38(2)*, 910–917.
- Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika 68(1)*, 165–176.
- Hurvich, C. M. and B. K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis 16*, 17–41.
- Jensen, M. J. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control 24(3)*, 361–387.
- Johnstone, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica 9*, 51–83.
- Johnstone, I. M. and B. W. Silverman (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B, Methodological 59*, 319–351.
- Lai, M.-J. (1995). On the digital filter associated with Daubechies’ wavelets. *IEEE Transactions on Signal Processing 43*, 2203–2205.
- Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7)*, 674–693.
- Masry, E. (1993). The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Transactions on Information Theory 39(1)*,

260–264.

- McCoy, E. and A. Walden (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* 5(1), 26–56.
- Nason, G. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society, Series B, Methodological* 58, 463–479.
- Ogden, R. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*. Boston: Birkhäuser.
- Percival, D. and A. Walden (1993). *Spectral Analysis for Physical Applications*. Cambridge, England: Cambridge University Press.
- Percival, D. and A. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press.
- Percival, D. B., S. Sardy, and A. C. Davison (2000). Wavestrapping time series: Adaptive wavelet-based bootstrapping. In W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing*, pp. 442–470. Cambridge, England: Cambridge University Press.
- Rangarajan, G. and M. Ding (Eds.) (2003). *Processes with Long-Range Correlations – Theory and Applications*. Lecture Notes in Physics. New York: Springer.
- Sabatini, A. M. (1999). Wavelet-based estimation of $1/f$ -type signal parameters: confidence intervals using the bootstrap. *IEEE Transactions on Signal Processing* 47(12), 3406–3409.
- Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York: Chapman & Hall.
- Strang, G. and T. Nguyen (1996). *Wavelet and Filter Banks*. Wellesley: Wellesley–Cambridge Press.
- Tewfik, A. and M. Kim (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory* 38, 904–909.
- Vannucci, M. and F. Corradi (1999). Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective. *Journal of the Royal Statistical Society, Series B, Methodological* 61, 971–986.

- Veitch, D. and P. Abry (1999). A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory* 45(3), 878–897.
- Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.
- Vidakovic, B. (1998). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.
- Yaglom, A. (1958). Correlation theory of processes with random stationary n th increments. *American Mathematical Society Translations (Series 2)* 8, 87–141.
- Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic results*. Springer-Verlag.
- Yajima, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *Journal of Time Series Analysis* 10, 375–383.

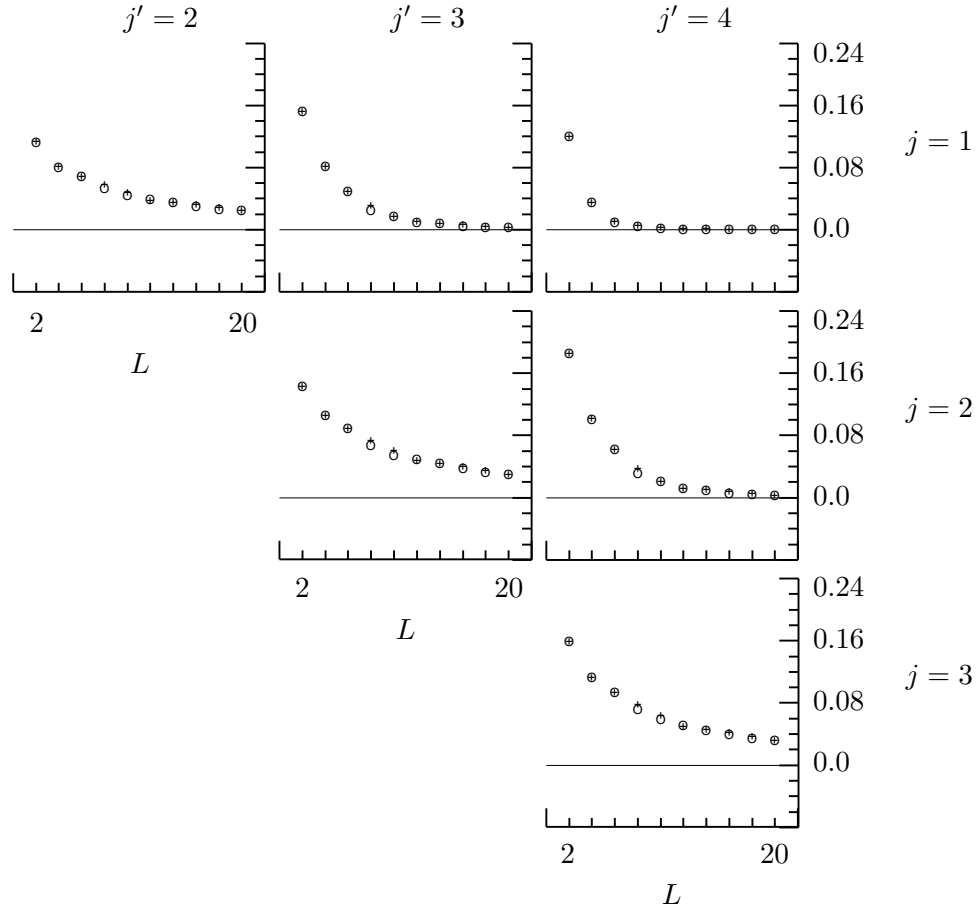


Figure 1: Maximum absolute correlation between wavelet coefficients $W_{j,t}$ and $W_{j',t'}$ for an FD process with $\delta = 0.4$. For $1 \leq j < j' \leq 4$, the maximum correlation was determined over all possible t and t' for Daubechies least asymmetric filters (circles) and extremal phase filters (pluses) of widths $L = 2, 4, \dots, 18$ and 20.

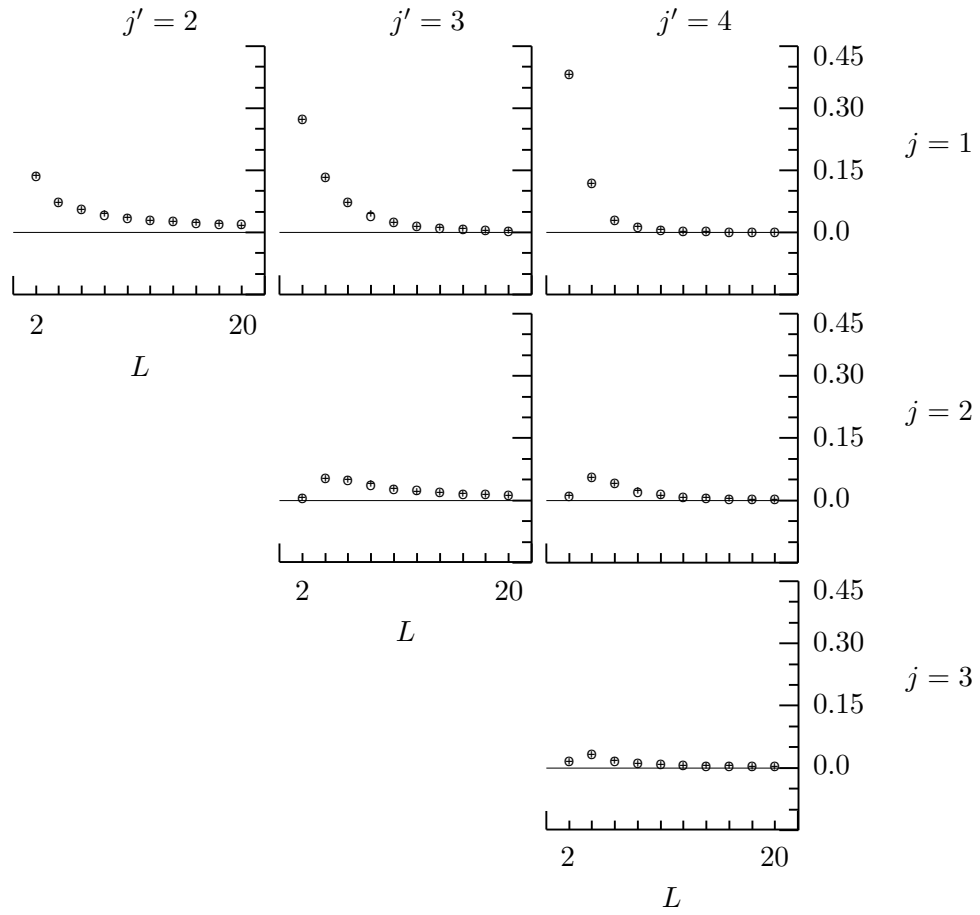


Figure 2: As in Figure 1, but now for an AR(1) process with $\phi = -0.9$.