Some New Statistical Approaches to the Analysis of Long Memory Processes

A thesis presented for the degree of Doctor of Philosophy
in the University of London
and the
Diploma of Membership of Imperial College
by

Emma J. McCoy

Department of Mathematics
Imperial College of Science Technology and Medicine
180 Queen's Gate, London SW7 2BZ

1994

Abstract

This thesis describes methods of analysis and synthesis of long memory processes. Long memory processes are those which exhibit correlations between events separated by a long period of time. This phenomenon is characterized in the frequency domain by a sharp peak in the spectral density function as the frequency approaches zero. This characteristic is observed in many physical time series, including those in the fields of geophysics, astronomy and finance.

A class of models that captures such long memory behaviour are fractionally differenced processes, the simplest of these processes is obtained by differencing white noise a fractional number of times.

We employ two methods of analyzing such processes: *Multitaper spectral estimation* and *Wavelet analysis*.

Multitaper spectral analysis uses the average of several direct spectral estimators evaluated using orthogonal tapers. We look at two sets of tapers: the *discrete prolate spheroidal sequences* and *sinusoidal tapers*. This method of spectral analysis reduces the variance of the estimator while maintaining protection against both local and broadband bias. We investigate properties of multitapering including the variance of multitaper spectral estimators and their effective bandwidth and show that multitapering provides improved estimators of the parameters of fractionally differenced processes over methods such as the periodogram or Burg estimators.

Wavelet analysis employs the discrete wavelet transform (DWT) of a time series. We look at the Haar wavelet and two families of wavelets created by Daubechies: the *least asymmetric wavelets* and the *extremal phase wavelets*. We show that the DWT of a time series exhibiting long memory has a simple covariance structure which can be exploited in the analysis and synthesis of such processes. We also investigate properties of the DWT including the impact of boundary effects caused by the finite data samples and the phase of the wavelet filters.

Acknowledgements

Firstly, I would like to thank my supervisor Andrew Walden for his excellent support, encouragement and generosity of time over the last three years. I would also like to thank Don Percival for giving me the opportunity to work at the University of Washington and for his continued help since my return.

Thanks also to all the members of the statistics section at Imperial college, especially Rodney Coleman for making many valuable comments on my transfer report.

I gratefully acknowledge the financial support of the Science and Engineering Research Council by way of a Research Studentship.

The continued support of close friends Colette and Volker provided a life beyond time series, and of course the unfailing support of my family, especially my sister Jacqui, for which I will always be grateful.

Finally, my greatest debt is to Stephen, for his continuing support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis introduces some new statistical techniques for the analysis of time series which display dependence between observations over long periods of time.

As a model for such time series we use *fractionally differenced processes*; which for certain parameterisations exhibit potential for modeling long memory behaviour.

We exploit the use of *multitaper spectral estimation* and the *discrete wavelet transform* to analyze such time series, and further investigate general properties of these techniques.

1.1 Outline of thesis

In this section we describe in detail the contents of the thesis.

Chapter 2 describes long memory processes and some of their determining characteristics and introduces a family of models – *fractionally differenced processes*, which arise as a generalisation of the well known ARMA models of Box & Jenkins (1976). These fractionally differenced models are referred to as ARFIMA models – fractionally integrated ARMA models. We show that these models have linear spectrum after suitable log transformations, a property which will enable us to easily estimate the long memory parameter of these models. A method of simulating time series from this class of models introduced by Hosking (1984) is also outlined.

Chapter 3 introduces the concept of *direct spectral estimation*, a method of spectral estimation in which we "taper" the data in order to guard against bias in the spectral estimator due to leakage – a common problem if the spectrum of interest has a large dynamic range. We introduce two families of orthogonal tapers; the *discrete prolate spheroidal sequences* and the *sinusoidal tapers*. We investigate the correlation of the estimators across frequencies and define a measure of bandwidth of the direct spectral estimators which can be used to measure the distance between approximately uncorrelated estimates.

Chapter 4 describes a method of spectral estimation known as *multitaper spectral estimation* which is formed from the average of several direct spectral estimators which have been calculated using orthogonal tapers. This method maintains the good bias properties associated with direct spectral estimation and at the same time produces an estimator with less variability. We establish the variance of multitaper spectral estimators for processes with a known autocovariance sequence, and show that over a large range of frequencies, as the number of tapers in the estimator increases so the variance decreases. We generalise the bandwidth measure derived for direct spectral estimators and show that it also provides an accurate measure of bandwidth for the multitaper estimators.

Chapter 5 demonstrates how multitaper estimators of the spectrum can be used to develop an estimator of the parameter of a fractionally differenced white noise process. We employ this method on both simulated series and some geological data, and compare the method to estimators obtained using previously published methods.

Chapter 6 introduces the *wavelet transform*, a technique which is becoming increasingly popular in signal and image processing as it decomposes the signal into components representing both scale and time. We describe some of the wavelets in current use and outline the algorithm used to obtain the *discrete wavelet transform*. We investigate boundary and phase effects of the transform. We look at the correlation structure of the transform coefficients, and in particular, demonstrate that for fractionally differenced white noise processes the discrete wavelet transform coefficients are approximately uncorrelated both across and within scales.

Chapter 7 examines how the discrete wavelet transform can be used to both synthesise and analyze long memory processes. We describe an algorithm that can be used to synthesise fractionally differenced white noise processes given their bandpass variances. We demonstrate that processes synthesised using this method have sample autocovariance sequences almost identical to processes synthesised using Hosking's method described in Chapter 2.

Finally, Chapter 8 summarises the research contained in this thesis and describes some possible directions for future research prompted from the work in this thesis.

Chapter 2

Long Memory Processes

2.1 Introduction

Many of the tools used in time series analysis depend upon the independence of observations separated by a long time span. More explicitly, these tools depend upon an exponential decay of the autocovariance sequence; many time series models have this property, including the well known ARMA models of Box & Jenkins (1976).

Many empirical time series, however, exhibit dependencies between distant observations. This long-term persistence can be characterized either by a non-exponential decay in the autocovariance sequence or alternatively, by a sharp increase in the spectral density as the frequency decreases towards zero. This dependence between distant observations can have a serious effect on classical tests and confidence intervals, most of which will have a slower rate of convergence, so any assumption of independence or short memory leads to underrating uncertainty.

The standard deviation of the sample mean of a long memory process can decay at a much slower rate than that of independent observations, implying that a much larger sample is required to achieve the same precision as in the more usual independent case.

Examples of such persistence can be found in many disciplines, including geophysics (Walden (1994)), hydrology (Hosking (1984)), astronomy (Deeter &

Boynton (1982), McHardy & Czerny (1987)) and meteorology (Haslett and Raftery (1989)). The first models for long memory behaviour were introduced by Mandelbrot & Van Ness (1968) in terms of fractional Brownian motions, and has since been extensively developed. The diversity of the applications means that work on the subject is scattered in the literature; for an excellent bibliographic guide of long memory processes see Taqqu (1986). Beran (1992b) gives an informative description of the history of long memory processes and a review of statistical methods for analyzing such processes.

2.2 Fractional differencing

In an attempt to model long memory behaviour Granger and Joyeux (1980) and Hosking (1981) introduced fractional ARIMA models, which are a generalisation of the standard ARIMA(p, d, q) models defined in Box & Jenkins (1976).

In the following we assume for simplicity that the sampling interval $\Delta t = 1$. When the differencing operator ∇ is applied to a time series $\{X_t\}$, d times, where d is a real number, a new series $\{\nabla^d X_t\}$ results. This series is defined by

$$\begin{aligned} \nabla^d X_t &= (1-B)^d X_t = \sum_{j=0}^{\infty} \begin{pmatrix} d \\ j \end{pmatrix} (-B)^j X_t \\ &= X_t - dX_{t-1} - \frac{1}{2}d(1-d)X_{t-2} - \frac{1}{6}d(1-d)(2-d)X_{t-3} - \ldots. \end{aligned}$$

If the derived times series is white noise $\{\epsilon_t\}$ with variance σ_ϵ^2, then $\{X_t\}$ is the simplest case of a fractionally differenced ARIMA process, a so called ARFIMA$(0, d, 0)$ process. It is this simplest type of fractional differencing that we investigate in this thesis, and refer to them as fractionally differenced processes, though strictly speaking they are fractionally differenced white noise processes.

Let $\{X_t\}$ be a zero mean ARFIMA$(0, d, 0)$ process with $-\frac{1}{2} < d < \frac{1}{2}$, then $\{X_t\}$ is both stationary and invertible. Given a sampling interval of Δt, the spectral density of $\{X_t\}$ is

$$S(f) = \Delta t \sigma_\epsilon^2 |2 \sin \pi f|^{-2d} \quad \text{for} \quad 0 < f < \frac{1}{2\Delta t}.$$

The autocovariance of $\{X_t\}$ is

$$s_\tau = E(X_t X_{t-\tau}) = \frac{\sigma_\epsilon^2 (-1)^\tau (-2d)!}{(\tau - d)!(-\tau - d)!}.$$

The variance of the process $\{X_t\}$ is

$$\mathrm{var}(X_t) = s_0 = \frac{\sigma_\epsilon^2 (-2d)!}{\{(-d)!\}^2}.$$

The partial correlations of $\{X_t\}$ are

$$\phi_{\tau\tau} = d/(\tau - d) \text{ for } \tau = 1, 2, \ldots. \quad (2.1)$$

For $d > 0$, both the decay of the correlations and the behaviour of the spectral density as the frequency decreases towards zero, indicate its potential for exhibiting long range persistence.

Figures 2.1 and 2.2 show respectively, the theoretical autocovariance sequence and spectrum for varying values of d. In Figure 2.3, we have also shown the spectrum versus both $\log(f)$ and $\log|2\sin(\pi f)|$ to show the linear relationships that exist between these variables.

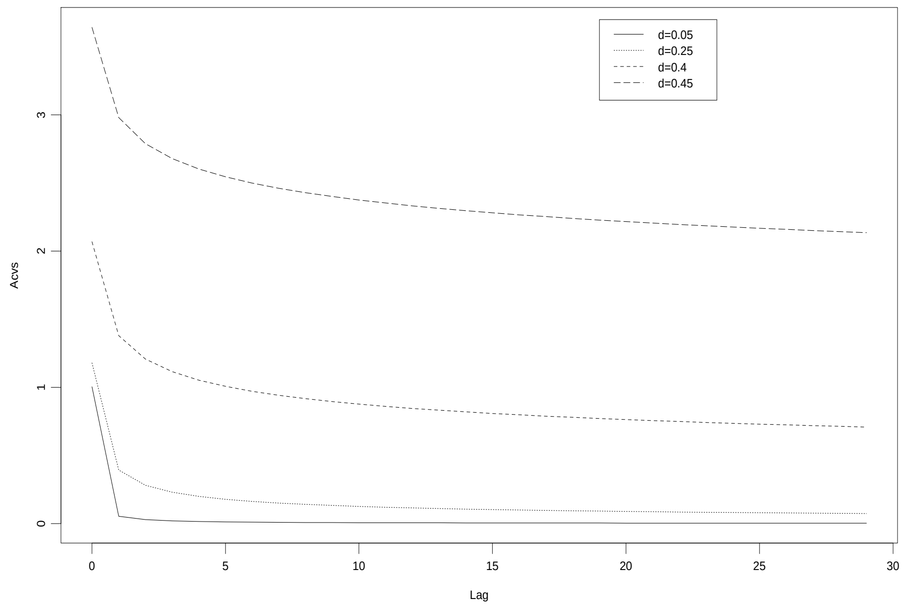


Figure 2.1: Autocovariance sequence for fractionally differenced processes

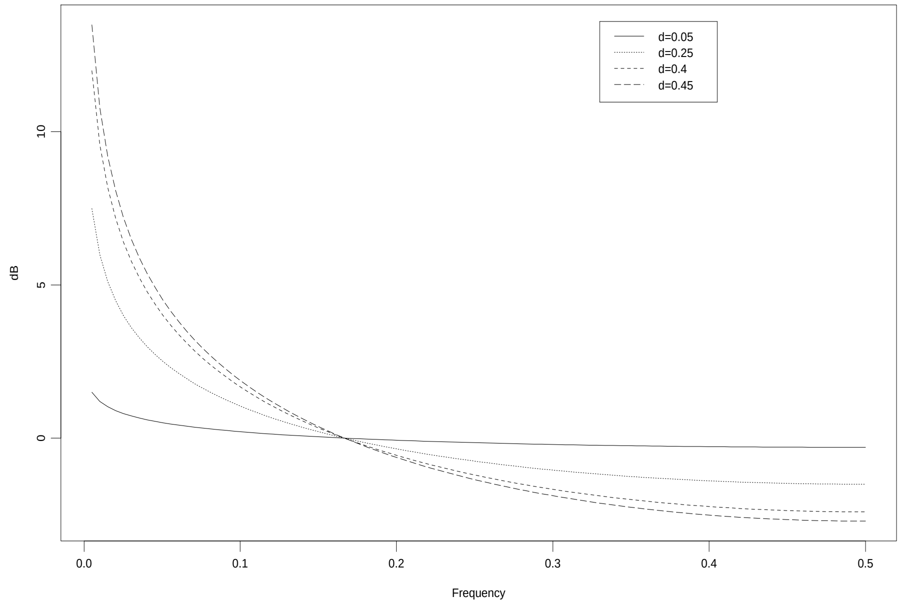


Figure 2.2: Spectrum for fractionally differenced processes

2.3 Simulating fractionally differenced processes

Here we describe a method of simulating ARFIMA$(0, d, 0)$ processes for d in the range $-\frac{1}{2} < d < \frac{1}{2}$ (Hosking (1984)).

For any stationary linear stochastic process $\{X_t\}$ with zero mean, the conditional mean and variance of $\{X_t\}$, given the past values $X_{t-1}, X_{t-2}, \ldots, X_0$ may be written

$$\begin{aligned} m_t &\equiv \mathrm{E}(X_t | X_{t-1}, X_{t-2}, \ldots, X_0) = \sum_{j=1}^{t} \phi_{tj} X_{t-j}, \\ v_t &\equiv \mathrm{var}(X_t | X_{t-1}, X_{t-2}, \ldots, X_0) = s_0 \prod_{j=1}^{t} (1 - \phi_{jj}^2) \end{aligned}$$

(Ramsey (1974)), where $s_0 = \mathrm{var}(X_t)$, ϕ_{ii} is the ith partial correlation coefficient of $\{X_t\}$ and the $\{\phi_{ti}\}$ are the partial linear regression coefficients, which can be calculated using the Levinson-Durbin recursion (Box & Jenkins (1976, p.84)).

An algorithm for simulating a sample $\{X_0, \ldots, X_{N-1}\}$ of size N from a stationary process with a normal marginal distribution and correlation function s_τ is described in Table 2.1.

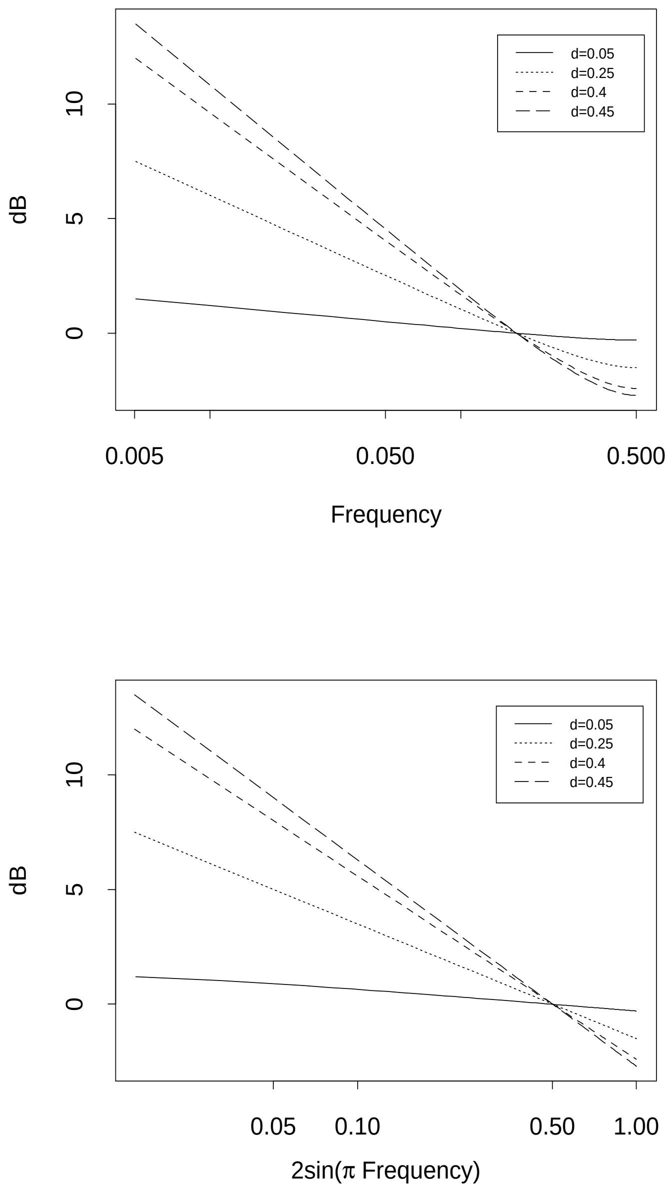


Figure 2.3: Transformed spectrum for fractionally differenced processes

1. Generate a starting value X_0 from the stationary distribution of the process $N(0, v_0)$, where v_0 $(= s_0)$ is the required variance of the $\{X_t\}$.

2. For $t = 1, \ldots, N-1$, calculate $\phi_{tj}, j = 1, \ldots, t$ recursively using the equations

$$\begin{aligned}
\phi_{tt} &= \frac{s_t - \sum_{j=1}^{t-1} \phi_{t-1,j} s_{t-j}}{v_{t-1}}, \\
\phi_{tj} &= \phi_{t-1,j} - \phi_{tt}\phi_{t-1,t-j} \text{ for } j = 1, \ldots, t-1.
\end{aligned}$$

3. Calculate $m_t = \sum_{j=1}^{t} \phi_{tj} X_{t-j}$ and $v_t = (1 - \phi_{tt}^2) v_{t-1}$.

4. Generate X_t from the distribution $N(m_t, v_t)$.

Table 2.1: Algorithm: Hosking's method for simulating fractionally differenced processes

This algorithm can be simplified by using the fact that the partial correlations for fractionally differenced white noise are given by Equation 2.1: $\phi_{tt} = d/(t-d)$.

Figure 2.4 shows simulations of fractionally differenced processes using the above algorithm. As the value of d increases the increase in the long term persistence becomes apparent.

2.4 Properties of long memory models

2.4.1 Measures of location and spread

Despite the fact that the correlations of a long memory process decay slowly, the sample mean of such a process does not lose much efficiency compared to the best linear unbiased estimator (see Samarov and Taqqu (1988) and Beran (1989,91)). This efficiency is very high for d in the range $[0, 1/2)$ (the stationary case), and so for most practical purposes the sample mean is a sufficiently accurate measure of location. Despite this efficiency, the standard deviation of the sample mean $\bar{X}_N$ of a long memory process of length N, decays at a rate of $N^{-a/2}$, $a \in (0,1)$, instead of $N^{-1/2}$, which is the usual case if the data are independent.

The classical scale estimator, the sample variance, is an inadequate estimator of the true variance. It can be badly biased, and loses much efficiency. Percival (1983) discusses the problem of scale estimation, and investigates alternative estimators of the variance.

Test for the presence of long memory can be found in Beran (1992a), who develops a goodness-of-fit test, and Agiakloglou & Newbold (1994), who develop a Lagrange multiplier test of ARMA(p, q) models against ARFIMA(p, d, q) alternatives.

In Chapter 5 we will be investigating frequency domain methods of estimating the parameter of fractionally differenced white noise – d. To do this we must first develop some spectral analysis techniques which will be outlined in the following two chapters.

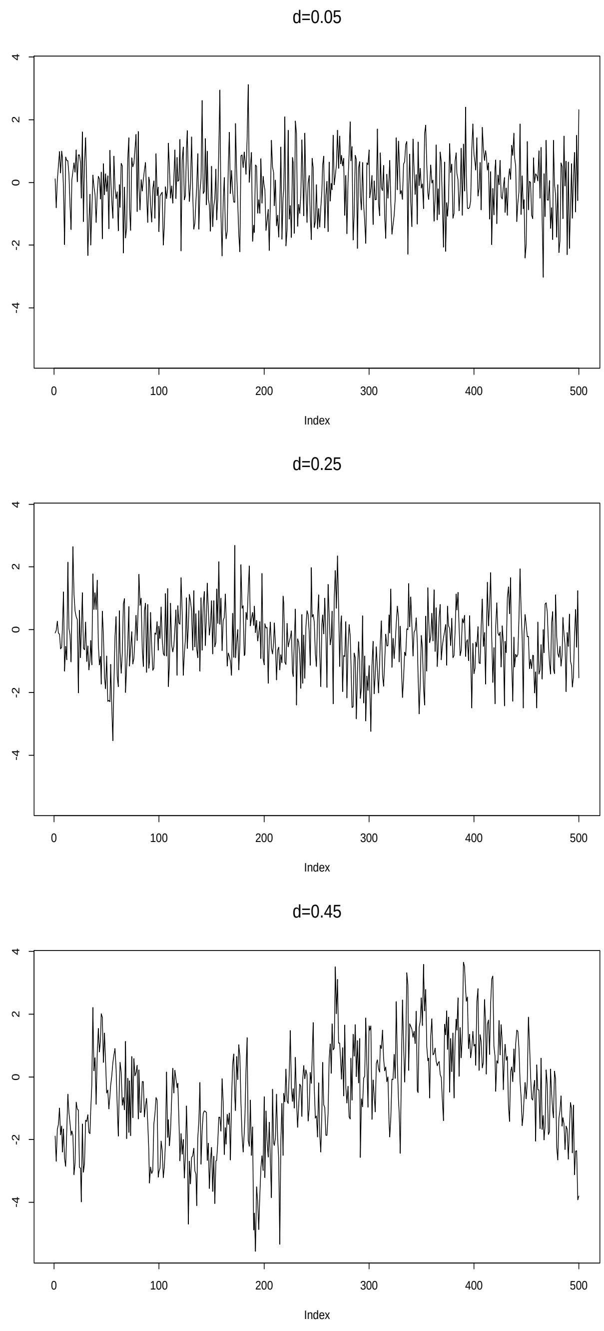


Figure 2.4: Simulated fractionally differenced processes

Chapter 3

Direct Spectral Estimation

3.1 Introduction

Spectral estimation is an invaluable tool in time series analysis, and is routinely employed in many of the physical sciences. Spectral analysis is, in effect, an analysis of variance technique which decomposes the variance of the time series into components at different frequencies. There are many methods of estimating the spectrum of a given time series, and some of these are outlined in Priestley (1981) and Percival & Walden (1993). In this chapter, we shall concentrate on a particular family of spectral estimators called *direct spectral estimators*.

New work contained in this chapter include the derivation of the exact form for the correlations between frequencies of a direct spectral estimator, and a new measure of bandwidth for direct spectral estimators. We also extend the work of Hurvich & Beltrao (1994) to calculate the asymptotic mean of a particular direct spectral estimator, which is used extensively in later chapters.

In order to motivate the use of direct spectral estimators, we shall first investigate a naïve estimator of the spectrum – the *periodogram*.

3.2 The periodogram

Given a time series $X_1, X_2, \ldots, X_N$, which we assume has zero mean and absolutely continuous spectral density function (sdf) given by $S(f)$, and autocovariance sequence (acvs) given by s_τ, then the following relationship between the spectrum and autocovariance sequence holds:

$$S(f) = \Delta t \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau\Delta t} \quad \text{with} \quad |f| \leq f_{(N)} \equiv \frac{1}{2\Delta t}, \tag{3.1}$$

where $f_{(N)}$ is the Nyquist frequency, and Δt is the sampling interval.

For long memory processes the acvs is not necessarily a square summable sequence, so that the relation 3.1 does not always hold. It is always the case that:

$$S^{(I)}(f_1, f_2) = \lim_{T\to\infty} \sum_{\tau=-T}^{T} \frac{s_\tau}{i\tau} \left(e^{-i2\pi f_1\tau\Delta t} - e^{-i2\pi f_2\tau\Delta t}\right), \tag{3.2}$$

where $S^{(I)}$ is the integrated spectrum (Yaglom (1987, p.183)).

A well known naïve estimator of the spectrum is given by the periodogram $\hat{S}^{(p)}(f)$, in which we substitute a biased estimator of the autocovariance, $\hat{s}_\tau^{(p)}$, in Equation 3.1:

$$\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} \text{ for } |\tau| \leq N-1, \tag{3.3}$$

giving

$$\begin{aligned}
\hat{S}^{(p)}(f) &= \Delta t \sum_{\tau=-(N-1)}^{(N-1)} \hat{s}_\tau^{(p)} e^{-i2\pi f\tau\Delta t} \\
&= \frac{\Delta t}{N} \sum_{\tau=-(N-1)}^{(N-1)} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} e^{-i2\pi f\tau\Delta t} \\
&= \frac{\Delta t}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} X_j X_k e^{-i2\pi f(k-j)\Delta t}.
\end{aligned}$$

Hence

$$\begin{aligned}
\hat{S}^{(p)}(f) &= \frac{\Delta t}{N} \sum_{j=1}^{N} X_j e^{i2\pi fj\Delta t} \sum_{k=1}^{N} X_k e^{-i2\pi fk\Delta t} \\
&= \frac{\Delta t}{N} \left|\sum_{k=1}^{N} X_k e^{-i2\pi fk\Delta t}\right|^2.
\end{aligned}$$

$\hat{S}^{(p)}(f)$ is defined over $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$.

The sequence $\{\hat{s}_\tau^{(p)}\}$ and $\hat{S}^{(p)}(f)$ are Fourier transform pairs, as are $\{s_\tau\}$ and $S(f)$. Note that we are truncating the summation over τ at the points $\pm(N-1)$; implicit in this truncation is the definition that $s_\tau = 0$ for $|\tau| > N$. While the periodogram can be a good approximation to the true spectrum in cases with small spectral range, in other cases it can be too severely biased to be useful.

In order to ascertain the bias of the periodogram we note the following:

$$\mathrm{E}\{\hat{S}^{(p)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \mathcal{F}(f - f')S(f')\,df', \tag{3.4}$$

where $\mathcal{F}(f)$ is Fejér's kernel, the spectral window of the periodogram:

$$\mathcal{F}(f) = \frac{\Delta t \sin^2(N\pi f \Delta t)}{N \sin^2(\pi f \Delta t)}.$$

Fejér's kernel is symmetric about the origin and consists of a broad central lobe of half-width $1/(N\Delta t)$, and $N-2$ sidelobes which decrease in magnitude as $|f|$ increases.

Figure 3.1 shows the shape of Fejér's kernel for $N = 32$ and $N = 128$, when $\Delta t = 1$, note the height of the sidelobes, and the fact that the main lobe has half-width $1/N$, marked by the solid vertical line.

As $N \to \infty$, Fejér's kernel acts as a Dirac δ function with an infinite spike at $f = 0$. Since $S(f)$ is assumed continuous, $\lim_{N\to\infty} \mathrm{E}\{\hat{S}^{(p)}(f)\} = S(f)$, and so the periodogram is an asymptotically unbiased estimator of $S(f)$.

Figure 3.2 shows the periodograms for the fractionally differenced datasets shown in Figure 2.4, the solid line shows the true spectrum, and we can see that as the value of d increases, the periodogram picks up the increase in power in the low frequency region.

We see from Equation 3.4 that the expected value of the periodogram is the convolution of the true spectrum with Fejér's kernel, so the shape of Fejér's kernel can cause bias in the periodogram as leakage occurs via the sidelobes.

The bias in a spectral estimator can be split into two distinct parts; the bias caused by loss of resolution within the central lobe: the local or smoothing bias, and bias caused by leakage via the sidelobes: the broadband bias.

One characterization of a given process is in terms of its *dynamic range*, this is defined by the ratio

$$10 \log_{10} \left(\frac{\max_f S(f)}{\min_f S(f)} \right).$$

The bias caused by Fejér's kernel is affected by the dynamic range of a process. Table 3.1 shows the dynamic range of fractionally differenced white noise over the

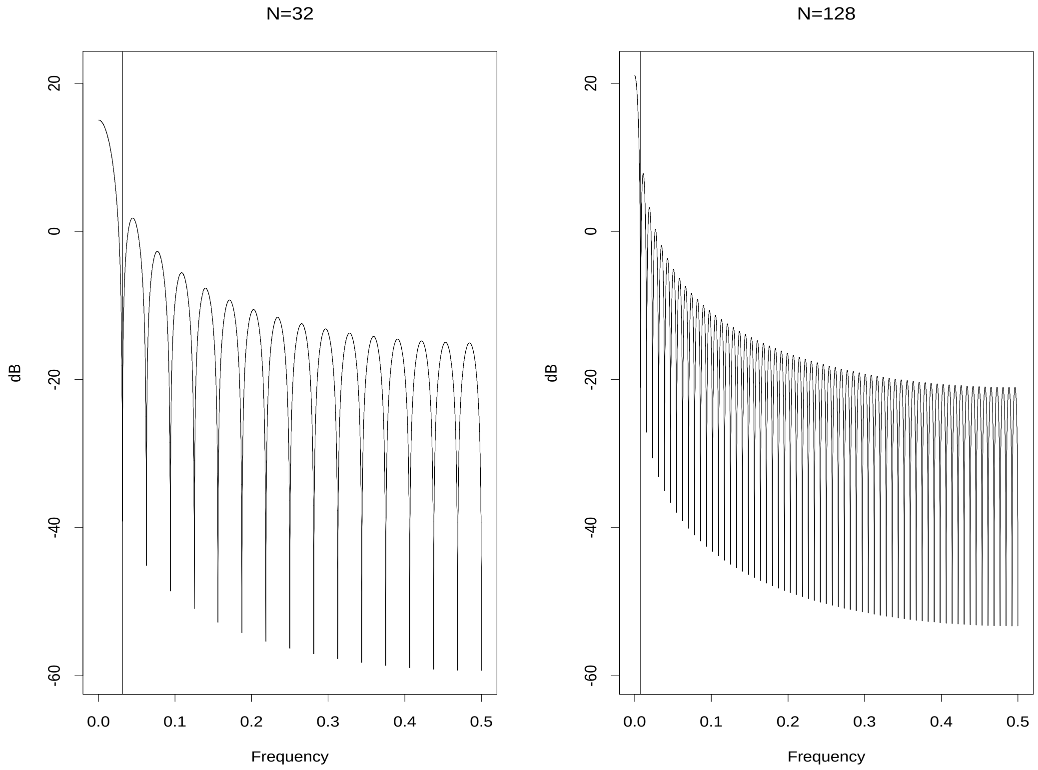


Figure 3.1: Fejér's kernel, $N = 32$ and $N = 128$

frequency range $1/N$ to $1/2$ (the range of the Fourier frequencies - $f_k = k/N\Delta t$, when $\Delta t = 1$), for several values of N and d; we see that as the value of d increases so the dynamic range becomes more severe.

Processes with a large dynamic range suffer from leakage to a greater extent than those with a lower dynamic range, especially when the spectrum is rapidly varying, as is the case for long memory processes.

One method of improving the bias properties of the periodogram is by tapering; this is where we modify Fejér's kernel to an alternative spectral window with lower sidelobes, thus decreasing the broadband bias.

3.3 Tapering

Let $X_1, X_2, \ldots, X_N$ be a section of a zero mean stationary process with sdf $S(f)$. We then form the product $h_t X_t$ for $1 \le t \le N$, where $\{h_t\}$ is called a data taper. We assume the convention that $\sum_{t=1}^{N} h_t^2 = 1$, which ensures the desirable property

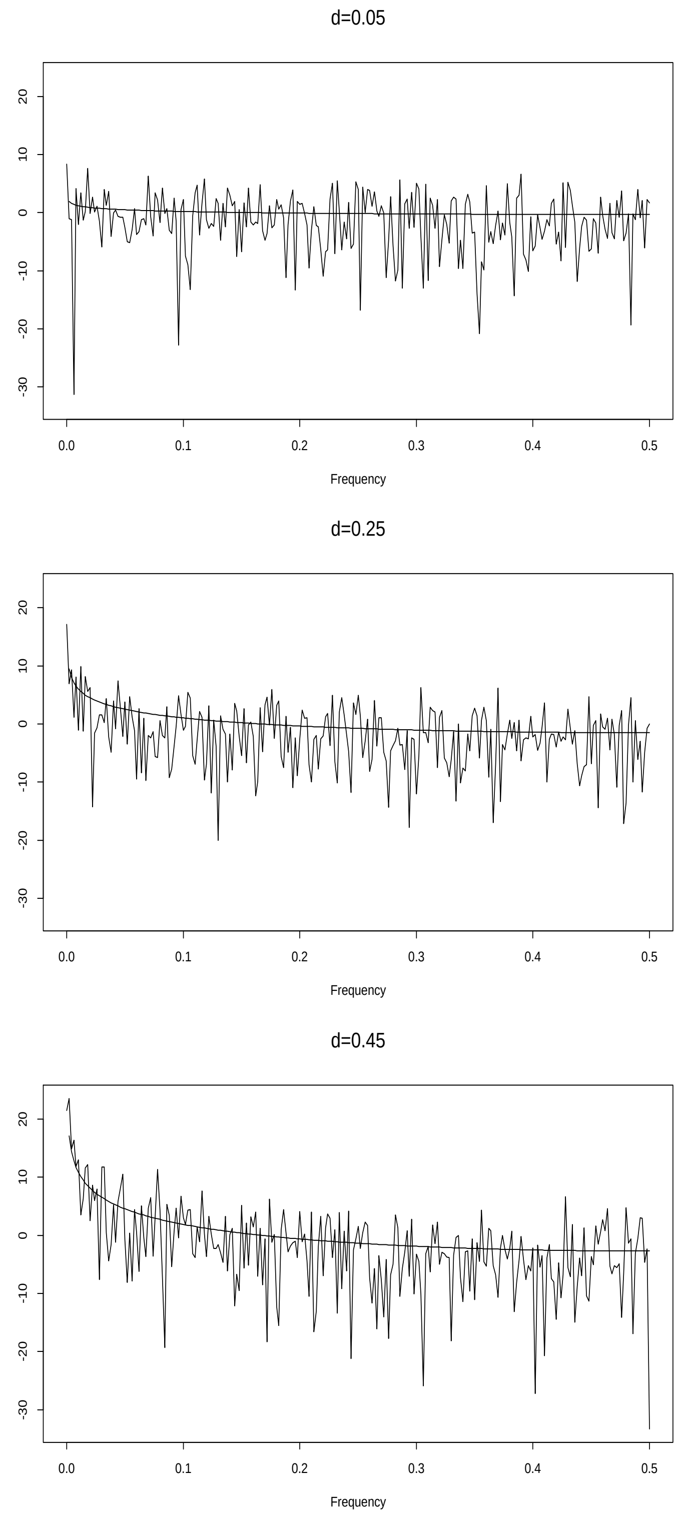


Figure 3.2: Periodogram's of simulated fractionally differenced processes

	Dynamic Range (dB)		
N	d=0.05	d=0.25	d=0.45
128	1.610	8.051	14.491
256	1.911	9.556	17.200
512	2.212	11.061	19.909
1024	2.513	12.566	22.618

Table 3.1: Dynamic range of fractionally differenced white noise (dB) in the frequency range $[1/N, 1/2]$, with evaluations at the Fourier frequencies $f_k = \frac{k}{N}$

that the resulting spectral estimator is unbiased for white noise. We can then calculate the direct spectral estimator, $\hat{S}^{(d)}(\cdot)$ using the tapered data,

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

We also note that

$$\mathrm{E}\{\hat{S}^{(d)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}(f - f')S(f')\, df',$$

where

$$\mathcal{H}(f) = \frac{1}{\Delta t}|H(f)|^2$$

and $H(f)$ is the Fourier transform of the data taper. That is:

$$H(f) = \Delta t \sum_{t=1}^{N} h_t e^{-i2\pi f t \Delta t},$$

$\mathcal{H}(f)$ is called the spectral window of the direct spectral estimator. If a rectangular data taper is used, i.e. $h_t = \frac{1}{\sqrt{N}}$ for $1 \leq t \leq N$, the direct spectral estimator reduces to the periodogram, and the spectral window $\mathcal{H}(f)$ becomes Fejér's kernel.

Our aim is to use a taper having a spectral window with superior properties to those of Fejér's kernel. In our case this implies employing a taper whose spectral window has low sidelobes to prevent leakage.

However, while tapering reduces bias due to leakage, there is a trade-off in that the effective size of the sample is reduced, resulting in a reduction in variance efficiency.

3.4 Orthogonal tapers

In this section we describe two sets of tapers which will be used extensively in our analyses; the *discrete prolate spheroidal sequences* and the *sinusoidal tapers.*

The first taper in the set of discrete prolate spheroidal sequences is called the zeroth order taper, and is designed to produce a direct spectral estimator with *minimum broadband bias*; the higher order discrete prolate spheroidal sequences ensure minimum broadband bias whilst being orthogonal to all lower order tapers.

The zeroth order sinusoidal taper produces a direct spectral estimator with *minimum local bias*, the higher order tapers ensure minimum local bias whilst being orthogonal to all lower order tapers.

3.4.1 Discrete prolate spheroidal sequences

In this section we outline the development of a series of orthogonal tapers which arise as a result of maximizing the energy of a function over a given frequency range. These tapers were first introduced by Slepian (1978).

Let $\{g_t\}$ be a real-valued sequence with finite energy and a sampling interval $\Delta t = 1$. The Fourier transform of $\{g_t\}$ is

$$G_p(f) = \sum_{t=-\infty}^{\infty} g_t e^{-i2\pi f t}. \tag{3.5}$$

Parseval's theorem states that

$$\sum_{t=-\infty}^{\infty} g_t^2 = \int_{-1/2}^{1/2} |G_p(f)|^2 \, df. \tag{3.6}$$

The energy in the frequency range $|f| \leq W < 1/2$ is $\int_{-W}^{W} |G_p(f)|^2 \, df$, and the fraction of energy in this range is

$$\beta^2(W) \equiv \frac{\int_{-W}^{W} |G_p(f)|^2 \, df}{\int_{-1/2}^{1/2} |G_p(f)|^2 \, df}. \tag{3.7}$$

Let us now consider the following concentration problem: to determine how large $\beta^2(W)$ in Equation 3.7 can be for $\{g_t\}$ index-limited to $[0, N-1]$. It follows from Equations 3.5 and 3.6 that Equation 3.7 can be rewritten as

$$\beta^2(W) = \frac{\sum_{t'=0}^{N-1} \sum_{t=0}^{N-1} g_t \frac{\sin[2\pi W(t'-t)]}{\pi(t'-t)} g_{t'}}{\sum_{t=0}^{N-1} g_t^2}. \tag{3.8}$$

To ascertain the sequence $\{g_t\}$ that maximizes $\beta^2(W)$ we can rewrite Equation 3.8 as

$$\beta^2(W) = \frac{\mathbf{g}^T A \mathbf{g}}{\mathbf{g}^T \mathbf{g}}, \tag{3.9}$$

where $\mathbf{g}$ denotes $g_0, g_1, \ldots, g_{N-1}$, and A denotes the $N \times N$ matrix with $A_{(t,t')} = \sin[2\pi W(t' - t)]/[\pi(t' - t)]$, the rows and columns of A are labeled from 0 to $N-1$. Differentiating Equation 3.9 with respect to $\mathbf{g}$:

$$\begin{aligned}
\frac{d}{d\mathbf{g}}\beta^2(W) &= \frac{\mathbf{g}^T\mathbf{g}\frac{d}{d\mathbf{g}}\left(\mathbf{g}^T A\mathbf{g}\right) - \mathbf{g}^T A\mathbf{g}\frac{d}{d\mathbf{g}}\left(\mathbf{g}^T\mathbf{g}\right)}{(\mathbf{g}^T\mathbf{g})^2} \\
&= \frac{\frac{d}{d\mathbf{g}}\left(\mathbf{g}^T A\mathbf{g}\right)}{\mathbf{g}^T\mathbf{g}} - \beta^2(W)\frac{\frac{d}{d\mathbf{g}}\left(\mathbf{g}^T\mathbf{g}\right)}{\mathbf{g}^T\mathbf{g}}.
\end{aligned}$$

But,

$$\frac{d}{d\mathbf{g}}\left(\mathbf{g}^T A\mathbf{g}\right) = 2A\mathbf{g} \quad \text{and} \quad \frac{d}{d\mathbf{g}}\left(\mathbf{g}^T\mathbf{g}\right) = 2\mathbf{g},$$

so that

$$\frac{d}{d\mathbf{g}}\beta^2(W) = 0 \Rightarrow 2\left[\frac{A\mathbf{g}}{\mathbf{g}^T\mathbf{g}} - \beta^2(W)\frac{\mathbf{g}}{\mathbf{g}^T\mathbf{g}}\right] = 0 \Rightarrow A\mathbf{g} = \beta^2(W)\mathbf{g}. \tag{3.10}$$

So the sequence $\{g_t\}$ that maximizes $\beta^2(W)$ must satisfy

$$A\mathbf{g} = \lambda_k(N, W)\mathbf{g}. \tag{3.11}$$

The Discrete Prolate Spheroidal Sequences (dpss's), are a series of index limited sequences, $\mathbf{v}_0(N, W), \ldots, \mathbf{v}_{N-1}(N, W)$, that correspond to the eigenvectors of Equation 3.11.

The zeroth order dpss $(v_{0,0}(N, W), \ldots, v_{N-1,0}(N, W))$ has the property that, of all index limited sequences, it has the highest concentration of energy in the user defined frequency interval $(-W, W)$. The first order dpss $(v_{0,1}(N, W), \ldots, v_{N-1,1}(N, W))$ has the highest concentration of energy in $(-W, W)$, with the condition that it is orthogonal to the zeroth order dpss, and so on, giving a collection of orthogonal tapers. From Equation 3.10, it follows immediately that the concentration of energy in the frequency band $(-W, W)$ of the kth taper is given by the kth eigenvalue – $\lambda_k(N, W)$.

Methods of calculating the discrete prolate spheroidal sequences can be found in Bell *et al* (1993), while an algorithm for calculating the zeroth order discrete prolate spheroidal sequence can be found in Walden (1989).

Figures 3.3 and 3.4 show both the tapers, $\mathbf{v}_k(N, W)$, and their corresponding spectral windows $\mathcal{H}_k$, for a variety of W and k, with $N = 32$, characterized by NW.

The fact that the majority of the energy is concentrated in a user specified bandwidth means that these tapers have good broadband bias properties. The width of the main lobe of the spectral window of the dpss's is larger than that of the periodogram, but the sidelobes are much lower.

The effectiveness of the dpss's as tapers decreases as their order increases, until a stage is reached where the majority of the energy is actually outside the frequency interval $(-W, W)$. The number of dpss's which are useful is dependent on the chosen bandwidth. In fact all orders of tapers less than the Shannon number (defined by $2NW\Delta t$), exhibit good leakage properties.

Figure 3.5 shows the direct spectral estimates of the data shown in Figure 2.4 using the zeroth order dpss as the taper, with $NW\Delta t = 4$, and $\Delta t = 1$. Comparing Figure 3.5 with Figure 3.2, which shows the periodograms of the same data, you can see that the direct spectral estimates vary more slowly than the periodograms. This is due to the change in width of the main lobe of the spectral window, which is wider for the dpss taper, resulting in a smoother estimate.

A single direct spectral estimator constructed from the higher order tapers would not be used as an estimator of the spectrum, as the zeroth order taper would always provide an improvement over higher order tapers in terms of their respective bias. Direct spectral estimators are, like the periodogram, inconsistent estimators, implying that their variance does not decrease as the sample size increases. We shall see in Chapter 4 that these orthogonal tapers can be used to calculate an estimator of the spectrum called a multitaper spectral estimator. This estimator has many attractive features, including reducing the variance of the estimator, while maintaining the good bias properties associated with tapering.

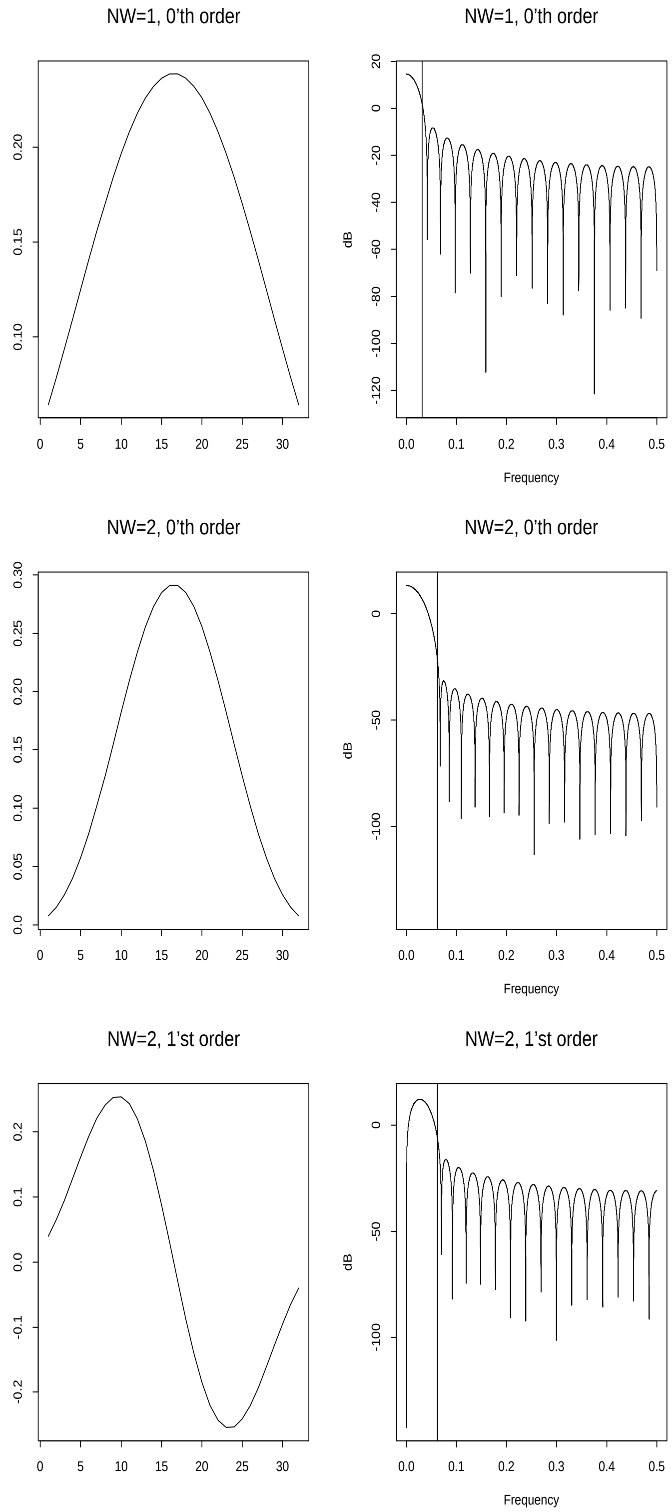


Figure 3.3: Discrete prolate spheroidal sequences and their associated spectral windows

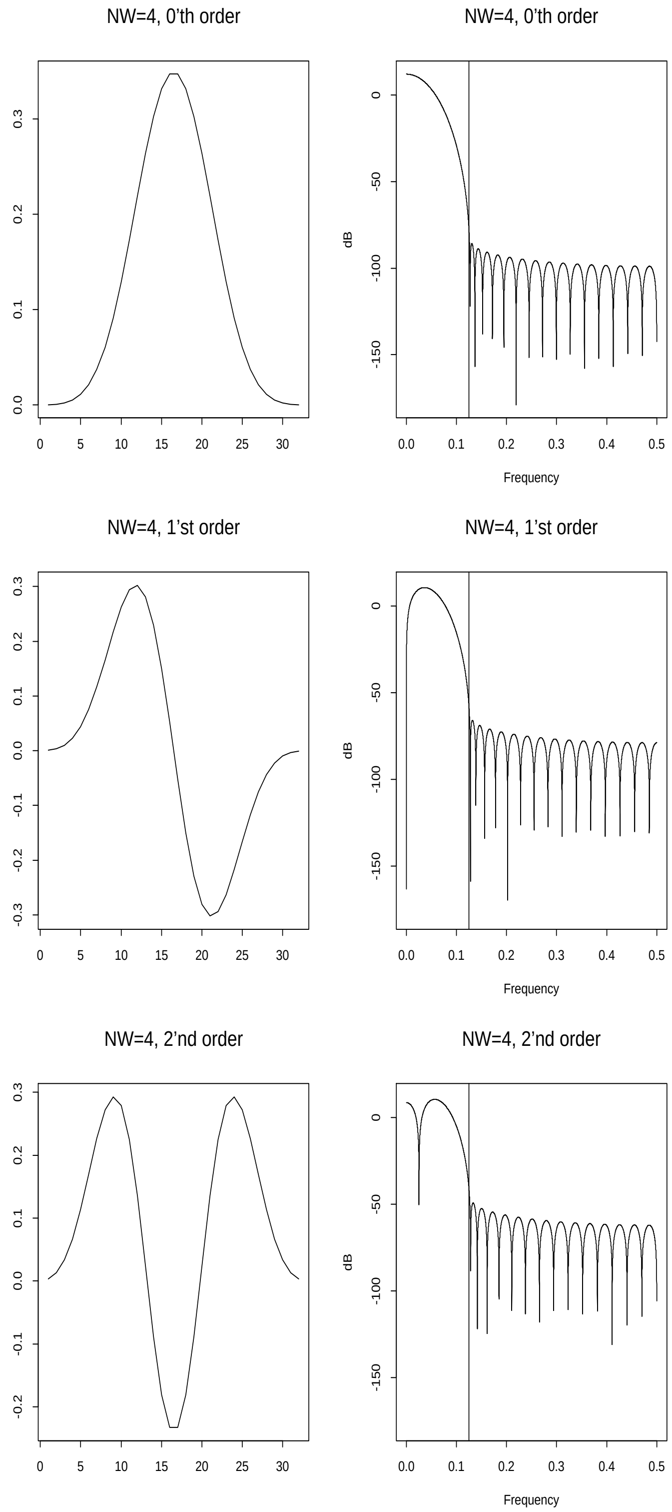


Figure 3.4: Discrete prolate spheroidal sequences and their associated spectral windows

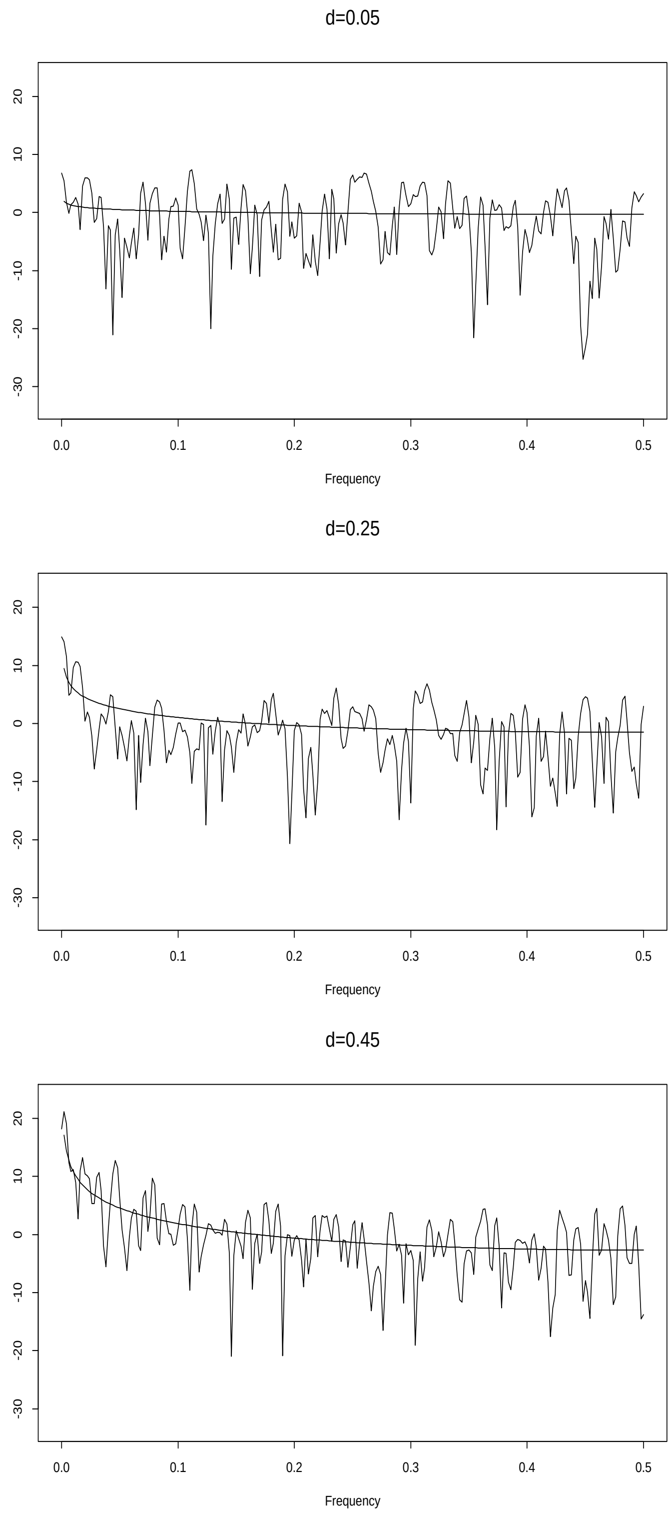


Figure 3.5: Direct spectral estimates for fdp's using zeroth order dpss taper

3.4.2 Sinusoidal tapers

In this section we describe a class of orthogonal tapers first introduced by Riedel & Sidorenko (1995). These tapers are designed to produce spectral estimators with minimum local bias.

Discrete minimum bias tapers

The local bias of a direct spectral estimator (assuming $f_{(N)} = 1/2$) is given by

$$E[\hat{S}(f)] - S(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(g-f)|^2 S(g)\, dg - S(f).$$

Using the facts that $|H(g-f)|^2$ and $S(f)$ are periodic and even, the fact that $\int_{-1/2}^{1/2} |H(f)|^2\, df = \sum_{t=0}^{N-1} h_t = 1$, making the substitution $h = g - f$ and using a Taylor series expansion for $S(\cdot)$ about f gives us the following form for the local bias:

$$\begin{aligned}
E[\hat{S}(f)] - S(f) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(g-f)|^2 \left(S(g) - S(f)\right) dg \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(h)|^2 \left(S(h+f) - S(f)\right) dh \\
&\approx \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(h)|^2 \left(S(f) + hS'(f) + \frac{h^2}{2} S''(f) - S(f)\right) dh \\
&= \frac{S''(f)}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(h)|^2 h^2\, dh.
\end{aligned}$$

For discrete tapers given by $\mathbf{h} = \{h_0, h_1, \ldots, h_{N-1}\}$, the leading order term of the local bias is proportional to

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 f^2 \mathrm{d}f &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| \sum_{n=0}^{N-1} h_n e^{-i2\pi f n} \right|^2 f^2\, df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} h_m h_n e^{-i2\pi f(m-n)} \right\} f^2\, df \\
&= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} h_m \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi f(m-n)} f^2\, df \right\} h_n \\
&= \mathbf{h} A \mathbf{h}^T,
\end{aligned}$$

where $A = [a_{nm}]$ with

$$a_{nm} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(n-m)f} f^2\, df = \begin{cases} \frac{1}{12} & \text{if } n = m \\ \frac{(-1)^{n-m}}{2\pi^2 (n-m)^2} & \text{if } n \neq m. \end{cases}$$

Following arguments similar to those used for the derivation of the discrete prolate spheroidal sequences given in Equations 3.9 to 3.11, it follows that the eigenvectors of A, sorted in increasing order of the eigenvalues, minimise the leading term in the local bias, subject to the constraints that the vectors are orthogonal, and the summability condition on the vectors holds ($\sum_{n=0}^{N-1} h_n^2 = 1$). The eigenvectors of A are called the *discrete minimum bias tapers*. The integral $\int_{-1/2}^{1/2} |H_k(f)|^2 f^2 \, df$, which is proportional to the local bias, is equal to the kth eigenvalue (recall that the concentration of energy in the defined bandwidth of the kth discrete prolate spheroidal sequence is the kth eigenvalue derived from Equation 3.11).

Discrete sinusoidal tapers

The discrete minimum bias tapers can be approximated by the *discrete sinusoidal tapers*, given by

$$h_{t,k} = \sqrt{\frac{2}{N+1}} \sin\left\{\frac{\pi(k+1)(t+1)}{N+1}\right\} \text{ for } k, t = 0, \ldots, N-1.$$

The square root term ensures that the resulting tapers are orthonormal.

Figures 3.6 and 3.7 show the minimum bias and discrete sinusoidal tapers and their spectral windows for $N = 32$ and $N = 128$. In the case where $N = 128$, the frequency axis has been split into frequencies close to zero and those close to 0.5 in order to obtain a clearer picture of the details.

Riedel & Sidorenko (1995) show that the the kth spectral window for the discrete sinusoidal tapers is concentrated on

$$\left[\frac{k}{2(N+1)}, \frac{k+2}{2(N+1)}\right] \bigcup \left[-\frac{k}{2(N+1)}, -\frac{k+2}{2(N+1)}\right].$$

The vertical lines on the plots of the spectral windows display these bounds.

From the figures one can see that in the time domain there is almost no difference between the minimum bias and the sinusoidal tapers. In the frequency domain there is little difference between the tapers at low frequencies, but at high frequencies the sinusoidal tapers are at a slightly higher level than the minimum bias tapers – indicating that in some cases the sinusoidal tapers may suffer from slightly more broadband bias due to leakage.

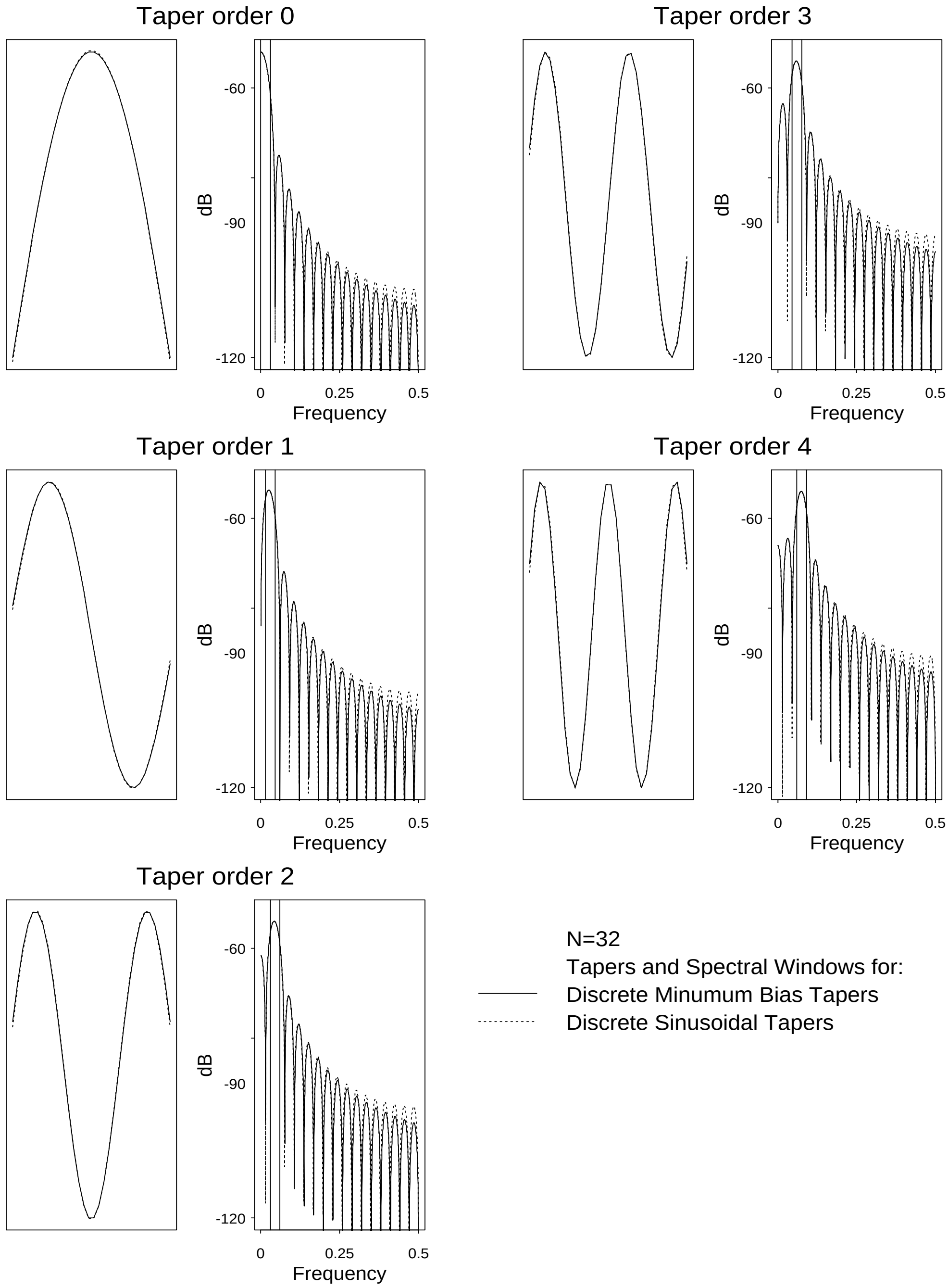


Figure 3.6: Discrete sinusoidal and minimum bias tapers and their spectral windows for $N = 32$

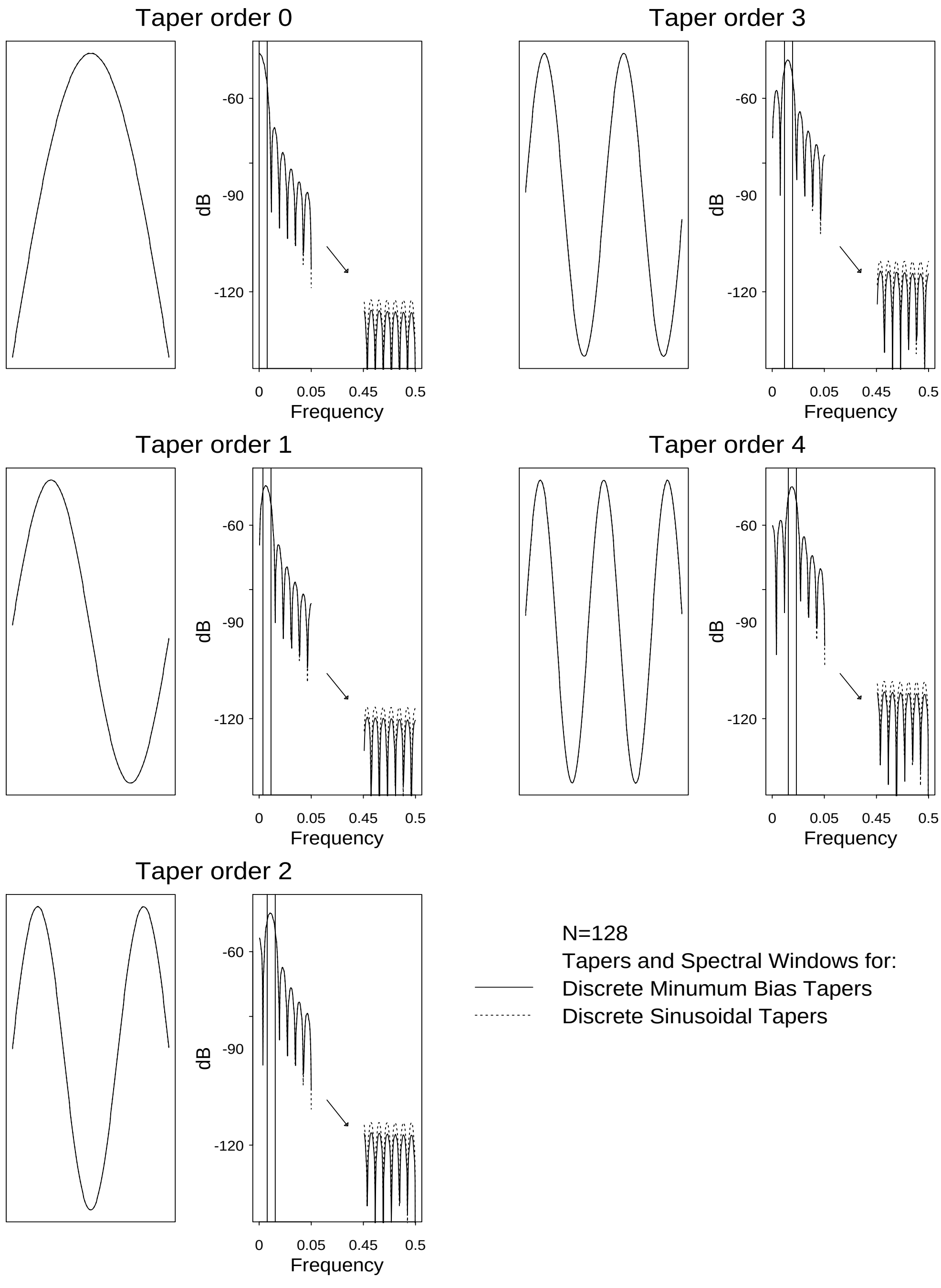


Figure 3.7: Discrete sinusoidal and minimum bias tapers and their spectral windows for $N = 128$

Unlike the discrete prolate spheroidal sequences, the discrete sinusoidal tapers are only controlled by one parameter – N, and the higher order tapers have a main lobe which is no longer centred on zero frequency. The usefulness of these higher order tapers will become apparent in the next chapter, where we introduce multitaper spectral estimation. For now, we shall just consider properties of direct spectral estimators constructed using the individual tapers.

By comparing Figures 3.6 and 3.7 with Figures 3.3 and 3.4 we can see that, for the chosen value of NW, the sidelobes of the discrete prolate spheroidal sequences tend to be lower than those of either the minimum bias tapers or the sinusoidal tapers. This should imply that direct spectral estimators constructed using discrete prolate spheroidal sequences will exhibit less broadband bias, however Riedel & Sidorenko (1995) argue that for most applications the minimum bias and sinusoidal tapers should provide more than adequate protection against both local and broadband bias.

3.5 Statistical properties of direct spectral estimators

Periodogram

For a zero mean stationary process $\{X_t\}$ with spectral density function $S(f)$ that is continuous over the Nyquist range, but that is not necessarily either white noise or Gaussian distributed, it can be shown – subject to the finiteness of certain high-order moments – that

$$\hat{S}^{(p)}(f) \stackrel{d}{=} \begin{cases} [S(f)/2]\chi_2^2, & \text{for } 0 < f < f_{(N)}; \\ S(f)\chi_1^2, & \text{for } f = 0 \text{ or } f_{(N)}, \end{cases}$$

as $N \to \infty$, where $\hat{S}^{(p)}(f)$ is the periodogram.

Furthermore, for $0 \le f' < f \le f_{(N)}$, $\hat{S}^{(p)}(f)$ and $\hat{S}^{(p)}(f')$ are asymptotically uncorrelated, giving

$$\operatorname{var}\{\hat{S}^{(p)}(f)\} \approx \begin{cases} S^2(f) & \text{for } 0 < f < f_{(N)}; \\ 2S^2(f) & \text{for } f = 0 \text{ or } f_{(N)}; \end{cases}$$

$$\text{cov}\{\hat{S}^{(p)}(f'), \hat{S}^{(p)}(f)\} \quad \approx \quad 0 \quad \text{for } 0 \leq f' < f \leq f_{(N)},$$

for large N.

Also, restricting ourselves to the Fourier frequencies $f_k = \frac{k}{N\Delta t}$, we find that the $\lfloor \frac{N}{2} \rfloor + 1$ rv's

$$\hat{S}^{(p)}(f_0), \hat{S}^{(p)}(f_1), \ldots, \hat{S}^{(p)}(f_{\lfloor \frac{N}{2} \rfloor})$$

are all approximately pairwise uncorrelated for large enough N.

Direct spectral estimators

The distributional theory for direct spectral estimators depend upon both the type of taper used and the true underlying sdf. For processes whose acvs satisfy a summability condition, and a taper of a reasonable form is used, Brillinger (1981, p.127) shows that the large sample distributional properties of direct spectral estimators are the same as those of the periodogram.

The grid of frequencies for which direct spectral estimators are pairwise uncorrelated is no longer the grid of Fourier frequencies, this is due to the change in the half-width of the central lobe from $\frac{1}{N\Delta t}$ in the case of the periodogram, to something larger in the case of a direct spectral estimator. For example, in the case of a dpss with $NW\Delta t = 2$, the grid size will be twice as large as the Fourier frequencies because the main lobe of the spectral window of the tapers is twice as wide as that of Fejér's kernel. Such widening of the main lobe causes local bias over a wider frequency band, thus affecting the correlation.

The limiting distributions of direct spectral estimators when the observations are drawn from a long memory process are no longer exactly those we have just described. These asymptotic distributions are discussed in Yajima (1989) and Hurvich & Beltrao (1993).

Yajima (1989) proves the asymptotic normality of the Fourier transform of a strongly dependent stationary process when evaluated at a set of fixed frequencies, not depending on the length of the series in question.

Hurvich & Beltrao (1993) investigate the asymptotic distribution of $\{\hat{S}^{(p)}(f_j)/S(f_j)\}_{j=1}^{M}$, where $\{f_j\}$ define the Fourier frequencies, for fixed M as $N \rightarrow \infty$, for a general long memory process. They show that this distribution depends upon both d and j, where d is the measure of persistence, and is an unequally weighted linear combination of independent χ_1^2 random variables.

These results, though apparently contradictory, are both correct. It is the different treatment of frequency which leads to the different results; Yajima considers a fixed frequency, as opposed to Hurvich & Beltrao who consider frequencies dependent on the length of the data through the definition of the Fourier frequencies.

Hurvich & Beltrao calculate the expected value of their limiting distribution for both the periodogram and a direct spectral estimator employing a cosine bell taper. In the Section 3.7 we shall follow their method to establish the asymptotic expected value of the lowest ordinates of the normalised direct spectral estimator employing the zeroth order discrete sinusoidal taper. We shall now investigate the differences between the cosine bell taper and the zeroth order discrete sinusoidal taper.

3.6 Comparison of tapers

The cosine bell taper is defined as follows:

$$\begin{aligned} h_t^{(cb)} &= \sqrt{\frac{2}{3N}}\left[1-\cos\left\{\frac{2\pi(t+1)}{N}\right\}\right] \quad \text{for} \quad t=0,1,\ldots,N-1 \\ &= \sqrt{\frac{8}{3N}}\sin^2\left\{\frac{\pi(t+1)}{N}\right\} = \sqrt{\frac{8}{3N}}\sin^2\{x_1\}, \end{aligned}$$

while the zeroth order discrete sinusoidal taper is defined as

$$\begin{aligned} h_t^{(s)} &= \sqrt{\frac{2}{N+1}}\sin\left\{\frac{\pi(t+1)}{N+1}\right\} \quad \text{for} \quad t=0,1,\ldots,N-1 \\ &= \sqrt{\frac{2}{N+1}}\sin\{x_2\}, \end{aligned}$$

with $x_1 = \pi(t+1)/N$ and $x_2 = \pi(t+1)/(N+1)$, and $x_1, x_2 \in (0,\pi)$. Now,

$$\begin{aligned} x_1 - x_2 &= \frac{\pi(t+1)}{N} - \frac{\pi(t+1)}{N+1} = \frac{\pi(t+1)}{N(N+1)} \\ &< \frac{\pi N}{N(N+1)} = \pi\left(\frac{1}{N+1}\right), \end{aligned}$$

which gives us the largest separation at where the sine or sine^2 term is evaluated between the two tapers. We deduce that the cosine bell taper has a sine^2 term evaluated at approximately the same value as the sine term in the discrete sinusoidal taper. This leads to the cosine bell taper being more concentrated around it's peak value, which occurs at the value closest of x_1 to $\pi/2$. Figure 3.8 show these tapers, and we see, as expected, that the cosine bell taper is more concentrated in time than the discrete sinusoidal taper. This concentration in time leads to a wider main lobe of the spectral window for the cosine bell taper, as is clearly seen in Figure 3.9.

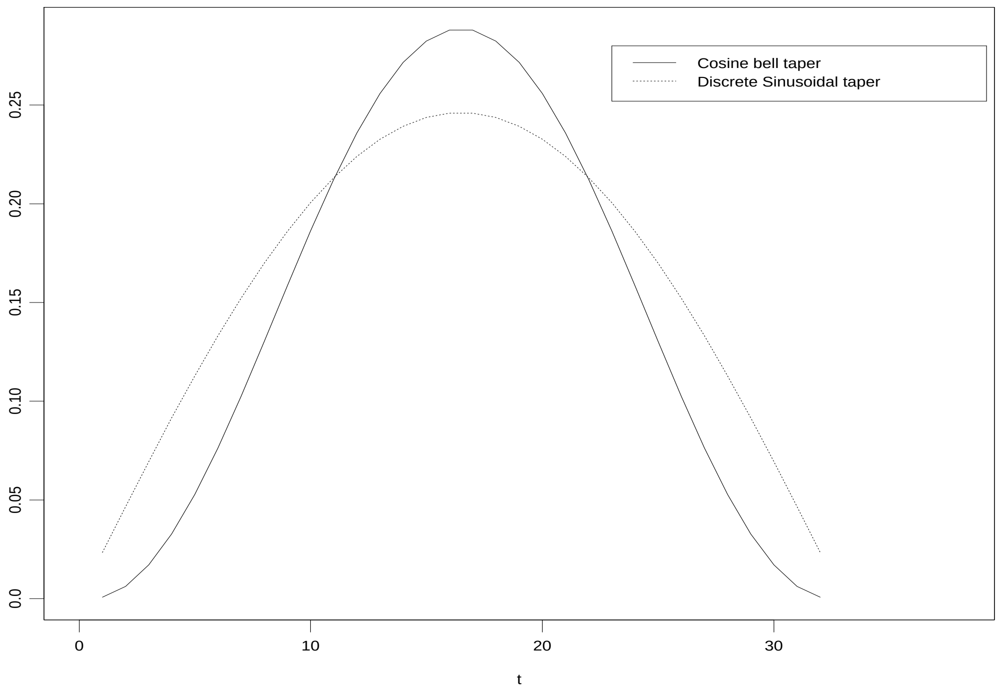


Figure 3.8: Comparison of tapers, $N = 32$

3.7 Asymptotic mean of normalized direct spectral estimators

In this section we establish the asymptotic mean of the normalized direct spectral estimator constructed using the zeroth order discrete sinusoidal taper. We

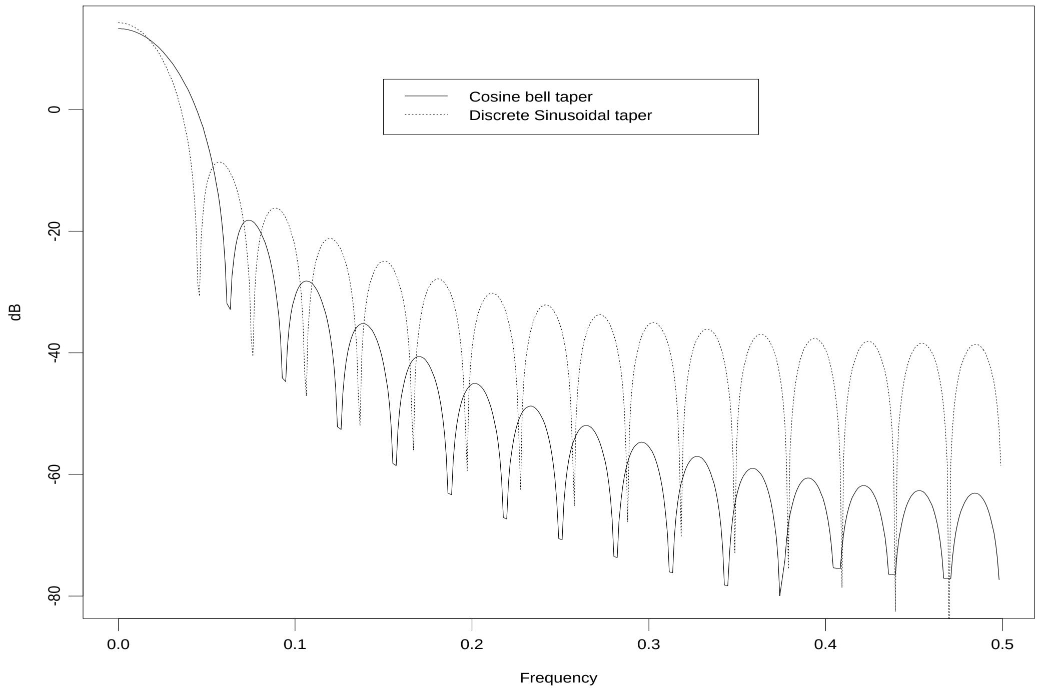


Figure 3.9: Comparison of spectral windows, $N = 32$

consider processes with zero mean and spectral density

$$S(f) = |1 - e^{-i2\pi f}|^{-2d} S^*(f) = |2\sin(\pi f)|^{-2d} S^*(f),$$

where $S^*(f)$ is an even positive function on $[-1/2, 1/2]$, bounded above and bounded away from zero. This class of densities include all fractionally differenced models. Without loss of generality, we can write the spectral density as $S(f) = |f|^{-2d} S^*(f)$, since both $|1-e^{-i2\pi f}|/|f|^{-2d}$ and $|f|^{-2d}/|1-e^{-i2\pi f}|$ are even, positive, continuous and bounded away from zero on $[-1/2, 1/2]$, and so can be incorporated into $S^*(f)$.

The spectral window of the zeroth order discrete sinusoidal taper is given by (see Riedel & Sidorenko (1995))

$$\mathcal{H}(f) = \frac{\sin^2\left(\frac{\pi}{N+1}\right)\cos^2\left([N+1]\pi f\right)}{2(N+1)\left\{\sin^2(\pi f) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2},$$

and the mean of $\hat{S}^{(d)}(f_j)/S(f_j)$ is given by

$$E\left\{\frac{\hat{S}^{(d)}(f_j)}{S(f_j)}\right\} = \frac{1}{S(f_j)}\int_{-1/2}^{1/2} \mathcal{H}(f_j - f')S(f')\,df'.$$

Define $g_N(f') = \mathcal{H}(f_j - f')S(f')/S(f_j)$. Then, letting $L_j(d)$ denote the limit of this mean, we have

$$L_j(d) = \lim_{N\to\infty} E\left\{\frac{\hat{S}^{(d)}(f_j)}{S(f_j)}\right\} = \lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} g_N(f')\, df'.$$

Yajima (1989) showed that at a fixed frequency f

$$\lim_{N\to\infty} E\left\{\frac{\hat{S}^{(d)}(f)}{S(f)}\right\} = 1$$

regardless of the value of d. We shall show that this is not the case for the lowest Fourier frequencies $f_j = j/N$ for small j.

Theorem

The asymptotic mean of the normalized direct spectral estimator of a fractionally differenced white noise process constructed using the zeroth order discrete sinusoidal taper is given by

$$L_j(d) = \frac{8}{\pi^2}\int_{-\infty}^{\infty} \frac{\cos^2(\pi\lambda)}{(4(j-\lambda)^2 - 1)^2}\left|\frac{\lambda}{j}\right|^{-2d}\, d\lambda.$$

Proof

Following Hurvich & Beltrao (1993) the proof of the theorem is split into two parts. In part 1 we establish that there exists $\alpha \in (0, 0.5]$ such that

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} g_N(f')\, df' = \lim_{N\to\infty} \int_{-N^{-\alpha}}^{N^{-\alpha}} g_N(f')\, df'. \tag{3.12}$$

In Part 2, we show that for the same α as in Equation 3.12,

$$\lim_{N\to\infty} \int_{-N^{-\alpha}}^{N^{-\alpha}} g_N(f')\, df' = \frac{8}{\pi^2}\int_{-\infty}^{\infty} \frac{\cos^2(\pi\lambda)}{(4(j-\lambda)^2 - 1)^2}\left|\frac{\lambda}{j}\right|^{-2d}\, d\lambda. \tag{3.13}$$

Here we shall only consider the case where $d > 0$, as it is for these values of d that the process exhibits long memory behaviour.

We now focus on part 1, the proof of Equation 3.12. First we show that there exists an $\alpha \in (0, 0.5]$ such that

$$\lim_{N\to\infty} \int_{N^{-\alpha}}^{\frac{1}{2}} g_N(f')\, df' = 0. \tag{3.14}$$

To establish Equation 3.14, since

$$\frac{S(f')}{S(f_j)} = \frac{S^*(f')|f'|^{-2d}}{S^*(f_j)(j/N)^{-2d}}$$

and since $S^*(\cdot)$ is bounded above and bounded away from zero, it suffices to show that

$$\lim_{N\to\infty} N^{-2d} \int_{N^{-\alpha}}^{\frac{1}{2}} \mathcal{H}(f_j - f')|f'|^{-2d}\, df' = 0. \tag{3.15}$$

The left hand side of Equation 3.15 is bounded by

$$\lim_{N\to\infty} \frac{N^{-2d}}{2} \sup_{f'\in[N^{-\alpha},\frac{1}{2}]} \mathcal{H}(f_j - f')|f'|^{-2d}.$$

Note that, for any $\alpha \in (0, 0.5]$, if $f' \in [N^{-\alpha}, 1/2]$ then

$$\begin{aligned} \mathcal{H}(f_j - f') &\leq \frac{1}{2(N+1)\left\{\sin^2(\pi[f_j - f']) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2} \\ &\leq \frac{1}{2(N+1)\left\{\sin^2(\pi[f_j - N^{-\alpha}]) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2}, \end{aligned}$$

for all sufficiently large N. To prove Equation 3.15 then, it suffices to show that there exists $\alpha \in (0, 0.5]$ such that

$$\lim_{N\to\infty} \frac{N^{-2d}}{2} \sup_{f'\in[N^{-\alpha},\frac{1}{2}]} \frac{|f'|^{-2d}}{2(N+1)\left\{\sin^2(\pi[f_j - N^{-\alpha}]) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2} = 0. \tag{3.16}$$

Put $\alpha = 0.25$. In this case,

$$\sup_{f'\in[N^{-\alpha},\frac{1}{2}]} |f'|^{-2d} = |N^{-0.25}|^{-2d} = N^{d/2}$$

(for sufficiently large N), and so the left-hand side of Equation 3.16 is bounded by

$$\begin{aligned} &\lim_{N\to\infty} \frac{N^{-2d}N^{d/2}}{4(N+1)\left\{\sin^2(\pi[f_j - N^{-0.25}]) - \sin^2\left(\frac{\pi}{[N+1]}\right)\right\}^2} \\ &= \lim_{N\to\infty} \frac{N^{-\frac{3d}{2}}}{4(N+1)\left\{(\pi(f_j - N^{-0.25}))^2 - \left(\frac{\pi}{[N+1]}\right)^2\right\}^2} \\ &= \lim_{N\to\infty} \frac{N^{-\frac{3d}{2}}}{4\pi^4(N+1)N^{-1}} \\ &= \frac{1}{4\pi^4} \lim_{N\to\infty} N^{-\frac{3d}{2}} = 0. \end{aligned}$$

Thus we have established Equation 3.16 and therefore Equations 3.15 and 3.14. A similar argument (omitted to save space) shows that, for the same value of α as used above, namely $\alpha = 0.25$, we have

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{-N^{-\alpha}} g_N(f')\, df' \;=\; 0.$$

It follows that, for α as given above, Equation 3.12 holds. Part 1 of the proof is complete.

For Part 2 of the proof, we need to establish Equation 3.13. Making the substitution $\lambda = Nf'$ we obtain

$$\begin{aligned}
\lim_{N\to\infty} \int_{-N^{-\alpha}}^{N^{-\alpha}} g_N(f')\,df' &= \lim_{N\to\infty} \int_{-N^{-\alpha}}^{N^{-\alpha}} \mathcal{H}(f_j - f') \frac{S(f')}{S(f_j)}\,df' \\
&= \lim_{N\to\infty} \int_{-N^{1-\alpha}}^{N^{1-\alpha}} \mathcal{H}\left(f_j - \frac{\lambda}{N}\right) S\left(\frac{\lambda}{N}\right) \frac{1}{S(f_j)} \frac{1}{N}\,df' \\
&= \lim_{N\to\infty} \left\{ \left(\frac{j}{N}\right)^{-2d} S^*(f_j) \right\}^{-1} \frac{1}{N} \int_{-N^{1-\alpha}}^{N^{1-\alpha}} \mathcal{H}\left(\frac{j-\lambda}{N}\right) \left|\frac{\lambda}{N}\right|^{-2d} S^*\left(\frac{\lambda}{N}\right) \mathrm{d}\lambda \\
&= \lim_{N\to\infty} \int_{-\infty}^{\infty} h_N(\lambda) \mathrm{d}\lambda,
\end{aligned}$$

where

$$h_N(\lambda) = \{S^*(f_j)\}^{-1} \frac{1}{N} \mathcal{H}\left(\frac{j-\lambda}{N}\right) \left|\frac{\lambda}{j}\right|^{-2d} S^*\left(\frac{\lambda}{N}\right) \chi_{[-N^{1-\alpha}, N^{1-\alpha}]}(\lambda),$$

and $\chi_{[-N^{1-\alpha}, N^{1-\alpha}]}(\lambda)$ is the indicator function of the interval $[-N^{1-\alpha}, N^{1-\alpha}]$. Defining $h(\lambda) = \lim_{N\to\infty} h_N(\lambda)$ and recalling that $0 < \alpha \le 0.5$ we have

$$\begin{aligned}
h(\lambda) &= \lim_{N\to\infty} \left\{ \frac{S^*\left(\frac{\lambda}{N}\right)}{S^*(f_j)} \frac{\sin^2\left(\frac{\pi}{N+1}\right) \cos^2\left([N+1]\pi\left[\frac{j-\lambda}{N}\right]\right)}{2N[N+1]\left\{\sin^2\left(\frac{\pi[j-\lambda]}{N}\right) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2} \left|\frac{\lambda}{j}\right|^{-2d} \right\} \\
&= \lim_{N\to\infty} \left\{ \frac{S^*\left(\frac{\lambda}{N}\right)}{S^*(f_j)} \frac{\frac{\pi^2}{[N+1]^2} \cos^2(\pi\lambda)}{2N[N+1]\left(\pi^2\frac{[j-\lambda]^2}{N^2} - \frac{\pi^2}{4[N+1]^2}\right)^2} \left|\frac{\lambda}{j}\right|^{-2d} \right\} \\
&= \lim_{N\to\infty} \left\{ \frac{16N^4\pi^2\cos^2(\pi\lambda)}{2N^4(4\pi^2(j-\lambda)^2 - \pi^2)^2} \left|\frac{\lambda}{j}\right|^{-2d} \right\} \\
&= \frac{8\cos^2(\pi\lambda)}{\pi^2(4(j-\lambda)^2 - 1)^2} \left|\frac{\lambda}{j}\right|^{-2d}.
\end{aligned}$$

We shall show that

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} h_N(\lambda)\,d\lambda = \int_{-\infty}^{\infty} h(\lambda)\,d\lambda, \tag{3.17}$$

thereby completing the proof of Equation 3.13. To establish Equation 3.17, we use Lebesgue's dominated convergence theorem. Let

$$C_1 = \sup_\lambda\{S^*(\lambda)\} / \inf_\lambda\{S^*(\lambda)\}.$$

There exists N_1 such that for all $N \geq N_1, j/N + N^{-\alpha} < 1/2$. Let $x = (j - \lambda)/N$. Then, for all $N \geq N_1$

$$
\begin{aligned}
\sup_\lambda \left\{ \frac{h_N(\lambda)}{h(\lambda)} \right\} &= C_1 \sup_{\lambda \in [-N^{1-\alpha}, N^{1-\alpha}]} \frac{\pi^2 \left[4(j-\lambda)^2 - 1\right]^2 \sin^2\left(\frac{\pi}{N+1}\right) \cos^2\left[(N+1)\pi\left(\frac{j-\lambda}{N}\right)\right]}{16N(N+1)\cos^2(\lambda\pi)\left\{\sin^2\left(\pi\left[\frac{j-\lambda}{N}\right]\right) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2} \\
&= C_1 \sup_{x \in [\frac{j}{N} - N^{-\alpha}, \frac{j}{N} + N^{-\alpha}]} \frac{\pi^2 \left[4N^2x^2 - 1\right]^2 \sin^2\left(\frac{\pi}{N+1}\right) \cos^2\left[(N+1)\pi x\right]}{16N(N+1)\cos^2([j - Nx]\pi)\left\{\sin^2(\pi x) - \sin^2\left(\frac{\pi}{2[N+1]}\right)\right\}^2} \\
&\approx C_1 \sup_{x \in [\frac{j}{N} - N^{-\alpha}, \frac{j}{N} + N^{-\alpha}]} \frac{\pi^2 \left[4N^2x^2 - 1\right]^2 \frac{\pi^2}{(N+1)^2} \cos^2\left[(N+1)\pi x\right]}{16N(N+1)\cos^2(Nx\pi)\left\{\sin^2(\pi x) - \frac{\pi^2}{4(N+1)^2}\right\}^2} \\
&\approx C_1 \sup_{x \in [\frac{j}{N} - N^{-\alpha}, \frac{j}{N} + N^{-\alpha}]} \frac{\pi^2 16N^4x^4 \frac{\pi^2}{N^2} \cos^2\left[(N+1)\pi x\right]}{16N(N+1)\cos^2(Nx\pi)\sin^4(\pi x)} \\
&\approx C_1 \sup_{x \in [\frac{j}{N} - N^{-\alpha}, \frac{j}{N} + N^{-\alpha}]} \frac{\pi^4 x^4}{\sin^4(\pi x)} \\
&= \frac{C_1 \pi^4 \left\{\frac{j}{N} + N^{-\alpha}\right\}^4}{\sin^4\left\{\pi\left(\frac{j}{N} + N^{-\alpha}\right)\right\}}, \qquad (3.18)
\end{aligned}
$$

for large enough N. Since the right-hand side of Equation 3.18 approaches C_1 as $N \to \infty$, there exists $N_2 \geq N_1$ such that for all $N \geq N_2$, $\sup_\lambda\{h_N(\lambda)/h(\lambda)\} \leq 1.01C_1$, and therefore for all λ, and for all $N \geq N_2$, $h_N(\lambda) \leq 1.01C_1h(\lambda)$. Since $1.01C_1h(\lambda)$ is integrable, Equation 3.17 follows by Lebesgue's dominated convergence theorem, and the proof is complete.

Hurvich & Beltrao (1993) give the following forms of $L_j(d)$ for the periodogram and the direct spectral estimator with the cosine bell taper.

Periodogram

The asymptotic mean of the normalized periodogram is given by

$$
L_{j,p}(d) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{\sin^2(\pi\lambda)}{(j-\lambda)^2} \left|\frac{\lambda}{j}\right|^{-2d} d\lambda.
$$

Tapered Periodogram

The asymptotic mean of the normalized direct spectral estimator constructed using the cosine bell taper is given by

$$\begin{aligned} L_{j,cb}(d) &= \frac{8}{3\pi^2}\int_{-\infty}^{\infty} \sin^2(\pi\lambda)\left[\frac{0.5}{j-\lambda} - \frac{0.25}{(j-1)-\lambda}\right. \\ &\qquad \left. - \frac{0.25}{(j+1)-\lambda}\right]^2 \left|\frac{\lambda}{j}\right|^{-2d} d\lambda. \end{aligned}$$

Figures 3.10 to 3.12 show these asymptotic means for $d \in [0, 0.5)$ and $j = 1, 2, \ldots, 10$. In agreement with Hurvich & Beltrao, we find that the periodogram has a positive asymptotic relative bias which diminishes slowly as j increases, however the bias is so small as to be almost inconsequential. Noticing the y-axis scale in Figure 3.12, we see that the maximum value of the asymptotic mean for the normalized periodogram is less than 1.15. In fact, Hurvich & Beltrao find that the asymptotic relative bias of the regression estimator of d (which we shall discuss in Chapter 5) based on the periodogram is only 3.92% when $d = 0.49$, which is a very severe value of d. They also examine d in the range $(-0.5, 0]$ and find that the bias is typically much greater for a given negative value of d than for the corresponding positive value $|d|$. This is not a surprising result if you consider that for negative d, the energy is concentrated in the high frequency part of the spectrum, and the periodogram offers little protection against broadband bias.

We also find that tapering reduces the asymptotic relative bias over the periodogram at all Fourier frequencies except the first, where tapering dramatically increases the asymptotic relative bias for d close to 0.5. This is probably due to the increase in width of the main lobe of the spectral window for the tapers over the spectral window for the periodogram. Notice that the discrete sinusoidal taper has less bias at the first Fourier frequency than the cosine bell taper, and again this is due to the difference in the widths of their respective spectral windows, as shown in Figure 3.9.

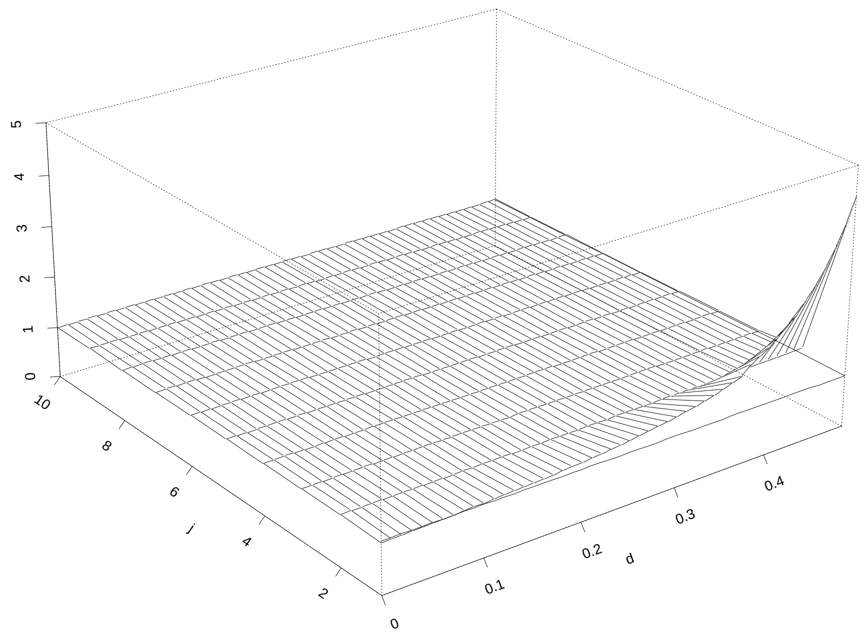


Figure 3.10: Asymptotic mean of normalized cosine direct spectral estimator

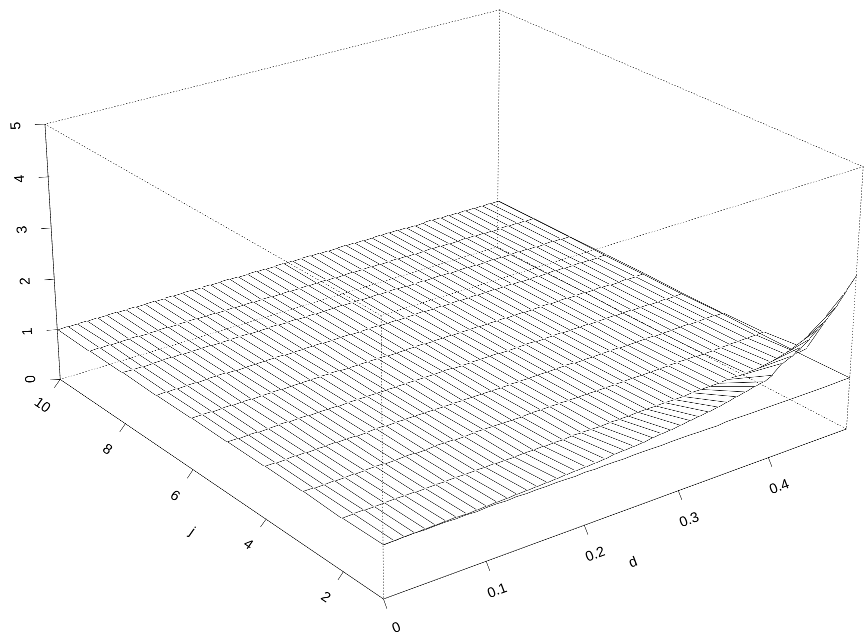


Figure 3.11: Asymptotic mean of normalized sine direct spectral estimator

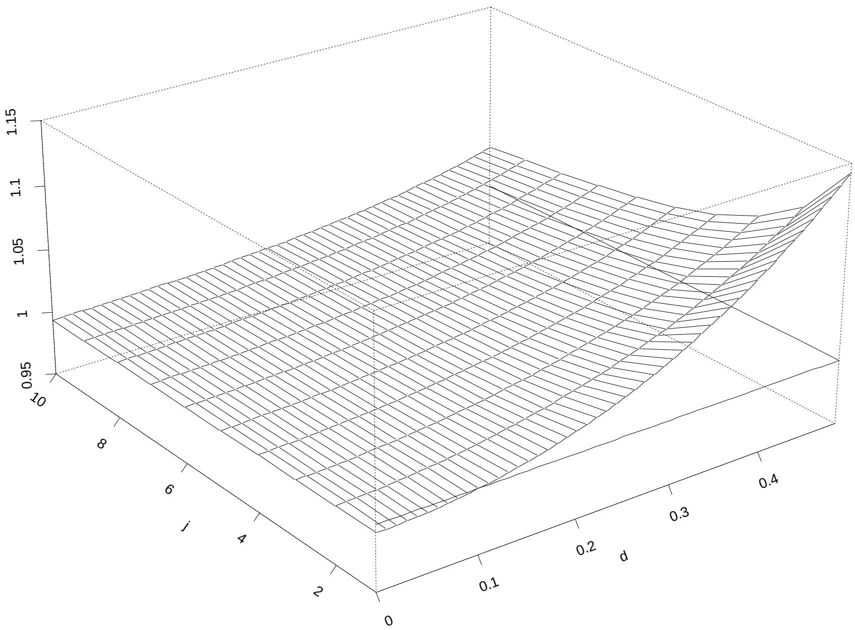


Figure 3.12: Asymptotic mean of normalized periodogram

3.8 Correlation structure of direct spectral estimators

As mentioned in the previous section, using a taper changes the grid of frequencies at which the estimates are pairwise uncorrelated. In this section we establish the exact form of the correlation between the spectral estimators at different frequencies for processes with a given acvs by applying a version of the Isserlis theorem. Some of this Section also appears in Walden, McCoy & Percival (1994).

Assuming that $\{X_t\}$ is a zero mean, Gaussian process. Let

$$d(f) = \sqrt{\Delta t} \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \Delta t}.$$

Then

$$\hat{S}^{(d)}(f) = d(f) d^*(f) \text{ and } \hat{S}^{(d)}(f') = d(f') d^*(f'),$$

where $*$ denotes the operation of complex conjugation. Note that $E\{d(f)\} =$

$E\{d^*(f)\} = 0$. We obtain the following equation for the covariance:

$$\text{cov}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f')\} = \text{cov}\{d(f)d^*(f), d(f')d^*(f')\}.$$

So, before we can proceed we need an expression for $\text{cov}\{Z_1Z_2, Z_3Z_4\}$ where (Z_1, Z_2, Z_3, Z_4) have the multivariate complex Gaussian distribution with zero mean and arbitrary covariance matrix. To clarify notation we firstly define

$$\text{cov}\{Z_j, Z_k\} = E\{Z_j^*Z_k\} - E\{Z_j^*\}E\{Z_k\},$$

so that

$$\text{cov}\{Z_1Z_2, Z_3Z_4\} = E\{Z_1^*Z_2^*Z_3Z_4\} - E\{Z_1^*Z_2^*\}E\{Z_3Z_4\}. \tag{3.19}$$

The complex version of the Isserlis theorem is:

$$E\{Z_1Z_2Z_3Z_4\} = E\{Z_1Z_2\}E\{Z_3Z_4\} + E\{Z_1Z_3\}E\{Z_2Z_4\} + E\{Z_1Z_4\}E\{Z_2Z_3\}.$$

Hence,

$$E\{Z_1^*Z_2^*Z_3Z_4\} = E\{Z_1^*Z_2^*\}E\{Z_3Z_4\} + E\{Z_1^*Z_3\}E\{Z_2^*Z_4\} + E\{Z_1^*Z_4\}E\{Z_2^*Z_3\}.$$

Thus, from Equation 3.19,

$$\text{cov}\{Z_1Z_2, Z_3Z_4\} = \text{cov}\{Z_1, Z_3\}\text{cov}\{Z_2, Z_4\} + \text{cov}\{Z_1, Z_4\}\text{cov}\{Z_2, Z_3\}. \tag{3.20}$$

It follows from Equation 3.20 that:

$$\begin{aligned}
\text{cov}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f')\} &= \text{cov}\{d(f)d^*(f), d(f')d^*(f')\} \\
&= \text{cov}\{d(f), d(f')\}\text{cov}\{d^*(f), d^*(f')\} + \\
&\qquad \text{cov}\{d(f), d^*(f')\}\text{cov}\{d^*(f), d(f')\} \\
&= \left|E\left\{d(f)d(f')\right\}\right|^2 + \left|E\left\{d(f)d^*(f')\right\}\right|^2 \\
&= (\Delta t)^2 E\left\{\left|\textstyle\sum_{u=1}^N h_u X_u e^{-i2\pi f' u\Delta t} \sum_{t=1}^N h_t X_t e^{-i2\pi f t\Delta t}\right|^2\right\} + \\
&\qquad (\Delta t)^2 E\left\{\left|\textstyle\sum_{u=1}^N h_u X_u e^{i2\pi f' u\Delta t} \sum_{t=1}^N h_t X_t e^{-i2\pi f t\Delta t}\right|^2\right\} \\
&= (\Delta t)^2 \left|\textstyle\sum_{u=1}^N h_u e^{-i2\pi f' u\Delta t} \sum_{t=1}^N h_t s_{t-u} e^{-i2\pi f t\Delta t}\right|^2 + \\
&\qquad (\Delta t)^2 \left|\textstyle\sum_{u=1}^N h_u e^{i2\pi f' u\Delta t} \sum_{t=1}^N h_t s_{t-u} e^{-i2\pi f t\Delta t}\right|^2.
\end{aligned} \tag{3.21}$$

Using this form of the covariance we can then calculate the correlation of the estimators at different frequencies:

$$\text{corr}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f')\} = \frac{\text{cov}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f')\}}{\sqrt{\text{var}\{\hat{S}^{(d)}(f)\}\text{var}\{\hat{S}^{(d)}(f')\}}}. \tag{3.22}$$

3.8.1 Correlation structure of the periodogram

The periodogram can be thought of as a direct spectral estimator employing a rectangular taper. The theory states that the periodogram is asymptotically uncorrelated at the Fourier frequencies for general $S(f)$. For example, setting $N = 32$, $h_t = 1/\sqrt{N}\ \forall\ t$, and $\Delta t = 1$, then substituting into Equations 3.21 and 3.22, we are in a position to investigate this claim for a variety of models.

White noise

As the dynamic range of white noise is zero we expect the periodogram to be well behaved in terms of broadband bias. For white noise, Equation 3.21 can be written

$$\operatorname{cov}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f+\eta)\} = \left\{ \left| \sum_{t=1}^{N} h_t^2 e^{i2\pi t\eta} \right|^2 + \left| \sum_{t=1}^{N} h_t^2 e^{-i2\pi(2ft+\eta t)} \right|^2 \right\},$$

and

$$\operatorname{var}\{\hat{S}^{(d)}(f)\} = \left\{ \left| \sum_{t=1}^{N} h_u^2 e^{-i4\pi ft} \right| + 1 \right\}.$$

which we can then substitute into Equation 3.22 to calculate the correlations.

Figure 3.13 shows these correlations, and we see that the periodogram is indeed uncorrelated at the grid of Fourier frequencies, shown by the vertical lines. The effect of the sidelobes of Fejér's kernel is clearly seen in the increase in correlation between the Fourier frequencies.

Fractionally differenced process

The shape of the spectrum of a fractionally differenced process may cause problems in terms of leakage in the spectral estimator. In fact, the periodogram still remains uncorrelated at the Fourier frequencies, even with a severe value of d. However, as can be seen from Figure 3.14, for frequencies close to zero, the periodogram is uncorrelated at the Fourier frequencies but becomes highly correlated at intermediate frequencies close to zero frequency. These correlations are due to the severe smoothing bias of the estimator within the main lobe.

As the frequencies increase, the shape of the spectrum at zero no longer affects the correlations, and they become comparable to the correlation for white noise.

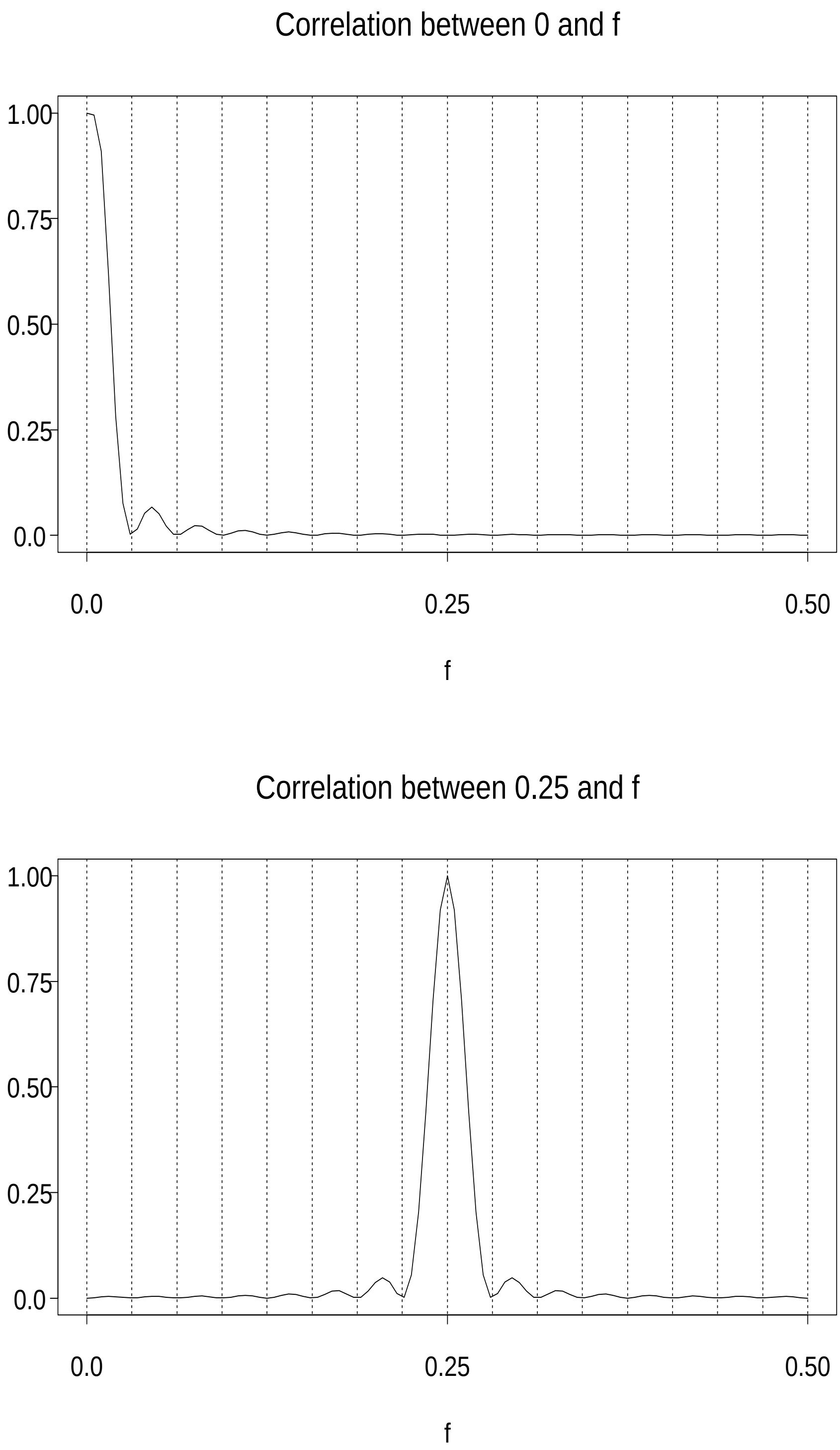


Figure 3.13: Correlations of periodogram for white noise

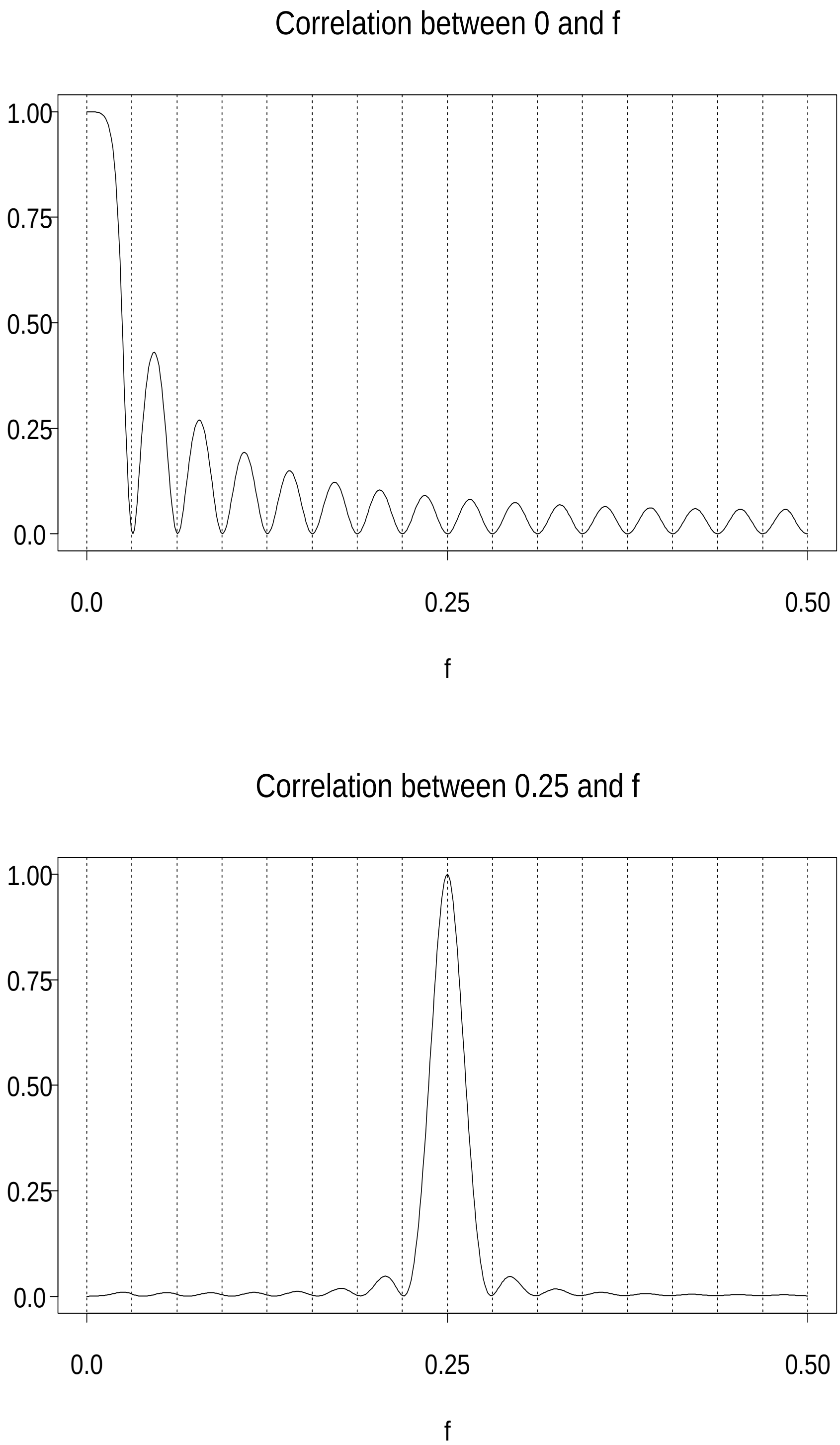


Figure 3.14: Correlations of periodogram for fdp(0.4)

3.8.2 Correlation structure of direct spectral estimators

Again, using Equations 3.21 and 3.22 we can investigate at what grid of frequencies direct spectral estimators are uncorrelated.

Using the zeroth to second order dpss tapers, and setting $NW\Delta t = 2$, with $N = 32$ and $\Delta t = 1$, we would expect the resulting direct spectral estimators to be approximately uncorrelated at a grid of frequencies twice as sparse as the Fourier frequencies.

We have also calculated the correlations using the zeroth to second order discrete sinusoidal tapers, with $N = 32$.

White noise

From the first row of Figure 3.15 it can be seen that the zeroth order taper behaves as is expected, with no correlation outside a half bandwidth on either side of the frequency in question, shown by the vertical lines. The higher order tapers, as can be seen from the second and third row, produce an estimator which exhibits surprising correlations at some frequencies.

Note that, unlike the periodogram, the estimators are also uncorrelated at intermediary frequencies outside $(f-W, f+W)$. This is due to the improvement in the broadband bias. As we would expect, for white noise, it does not matter at which frequency we are measuring correlation, this can be seen by noting the similarity in shape of the correlations between the first and second column of Figure 3.15.

From Figure 3.16 we see that, similarly to the discrete prolate spheroidal sequences, the higher order sinusoidal tapers give rise to some unexpected correlations. The vertical lines in this figure show a nominal half bandwidth on either side of the frequency in question. Recall from Section 3.4.2, that the nominal half bandwidth of the kth order discrete sinusoidal taper is given by $1/(N+1)$.

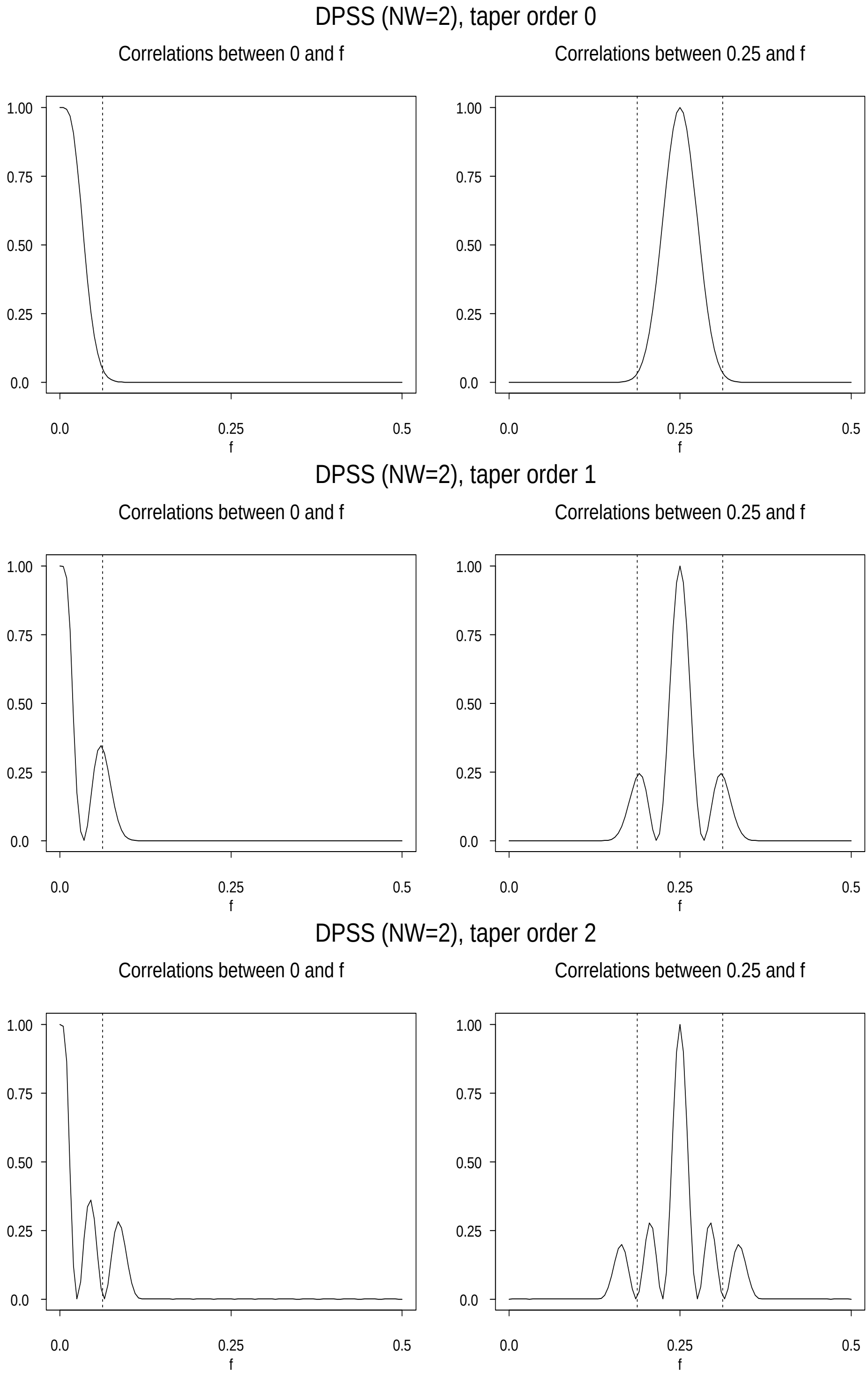


Figure 3.15: Correlations of direct spectral estimators for white noise

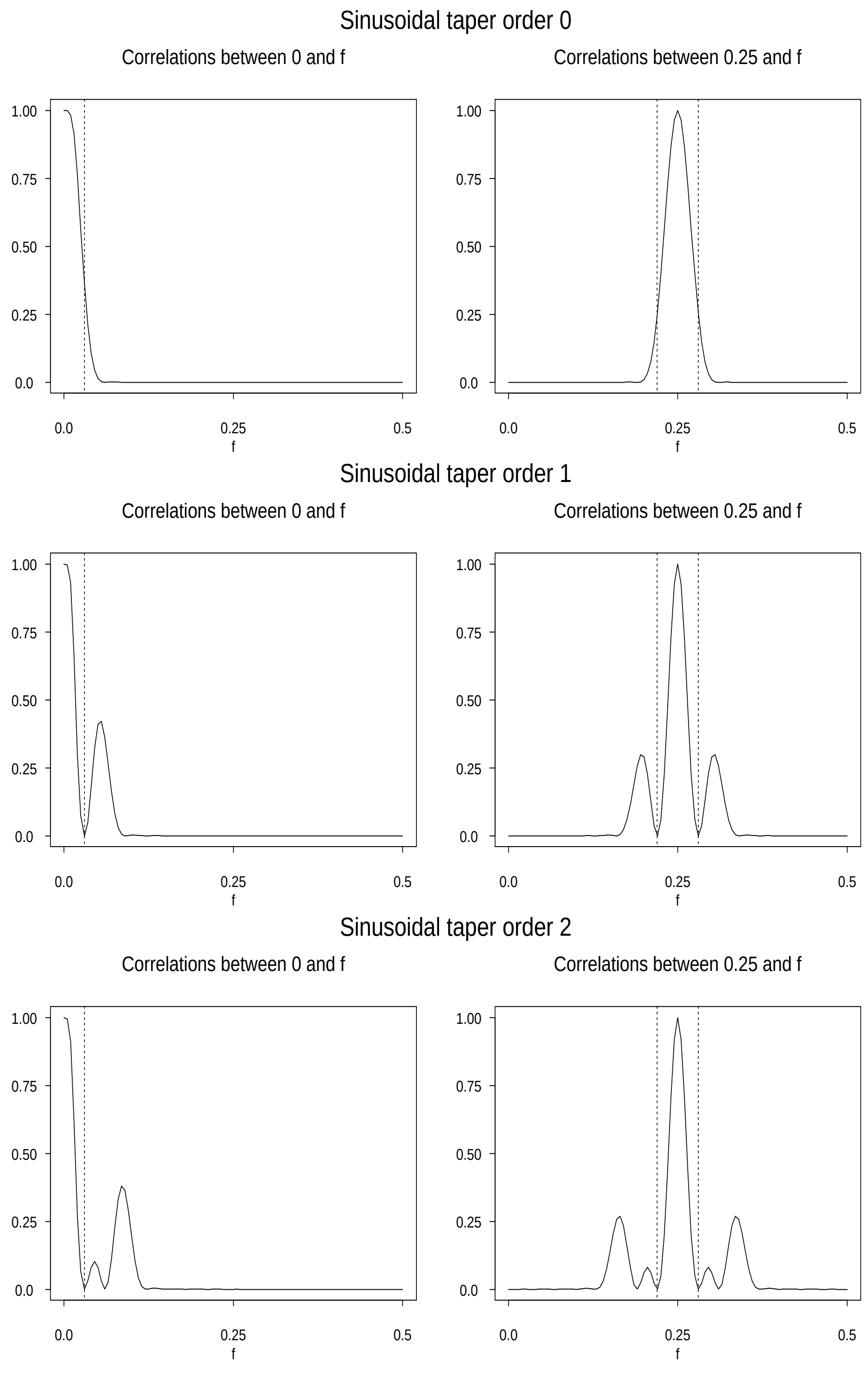


Figure 3.16: Correlations of direct spectral estimators for white noise

Fractionally differenced process

For a fractionally differenced process with $d = 0.4$, we see from the first row of Figure 3.17, that the zeroth order taper is well behaved. The second and third rows of the first column, show that, for frequencies close to zero, the higher order tapers exhibit correlation at some frequencies where it is theoretically uncorrelated. However, as the estimator moves away from zero frequency, shown in the second column, then the correlations begin to behave as in the white noise case.

We can see a similar pattern of behaviour in Figure 3.18, which shows the correlations for the discrete sinusoidal tapers.

3.8.3 Implications of correlation

In later chapters we will be outlining approaches to estimating d, the measure of persistence of a fractionally differenced process. Some of these methods use regression of the log spectrum to estimate d, and it is important that we have an understanding of the correlation structure of any estimator of the spectrum before we perform any regression techniques.

If we are aware of the correlations between frequencies, then this enables us to adjust our regression accordingly to take these correlations into account. This method of adjusting the regression is explained in more detail in Section 5.2, and employs a generalisation of the measure of bandwidth that we describe in the following section.

3.9 Bandwidth of direct spectral estimators

One method of establishing the grid of frequencies at which any given spectral estimator is uncorrelated, is by employing a bandwidth measure. The bandwidth of a spectral estimator is a measure of the minimum separation in frequency between approximately uncorrelated spectral estimators. We will utilise such a measure in Section 5.2 to decide which points to use in a regression of the estimated spectrum in order to calculate the measure of persistence for long memory models.

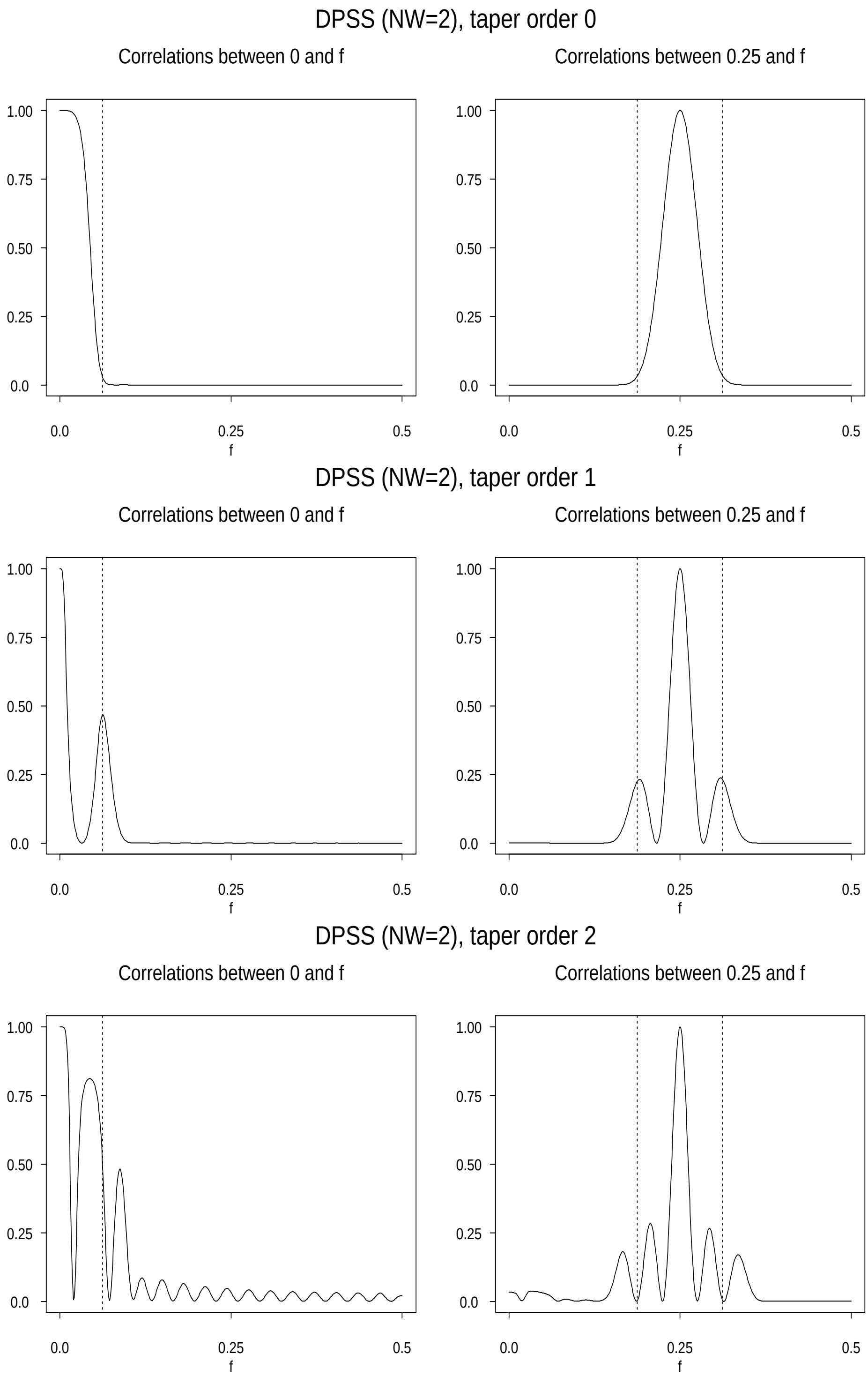


Figure 3.17: Correlations of direct spectral estimators for fdp(0.4)

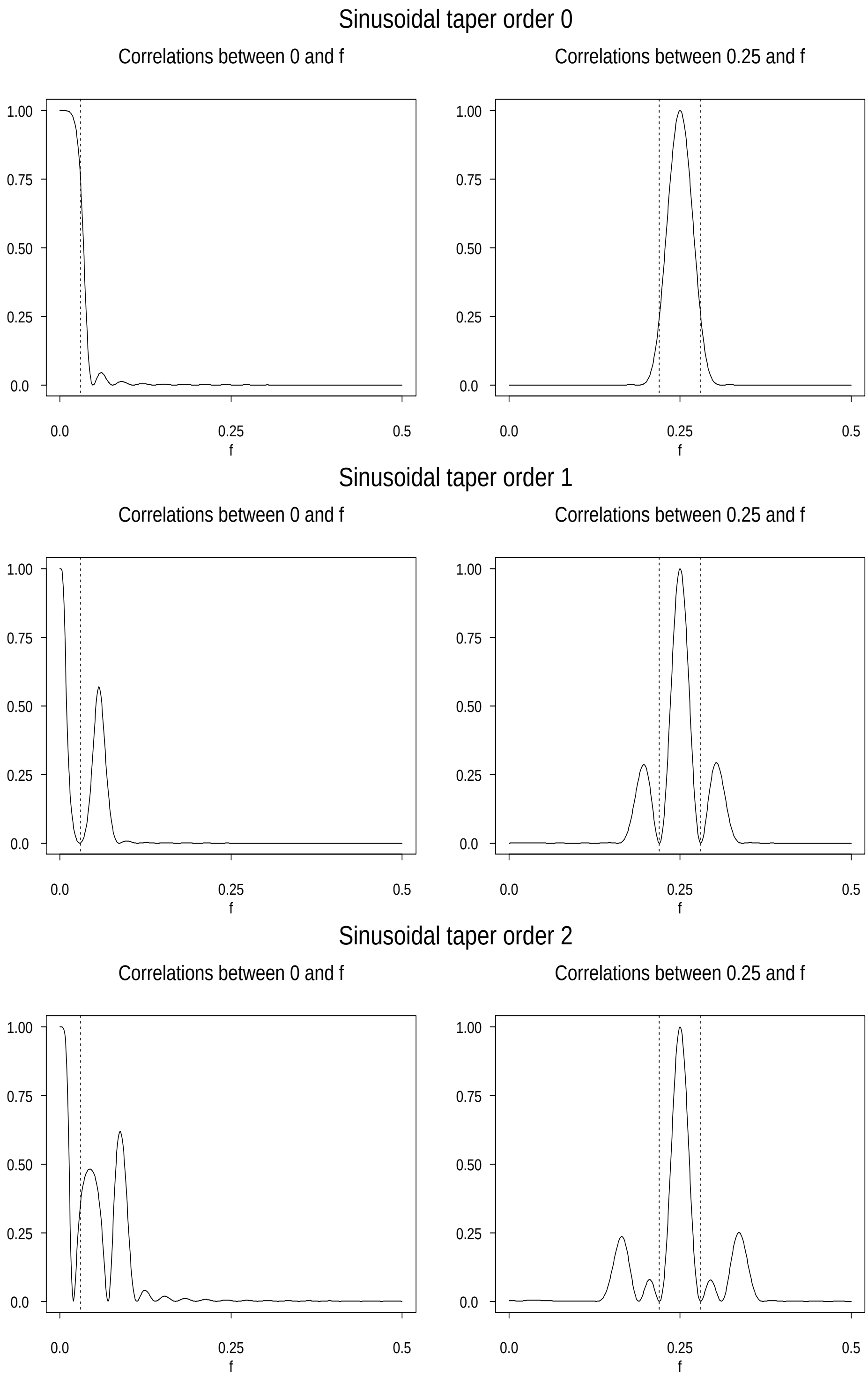


Figure 3.18: Correlations of direct spectral estimators for fdp(0.4)

Another important application of bandwidth is in comparisons between two different spectral analysis methods, which should ideally be based upon using the same bandwidth in each of the methods.

In this section we shall consider a measure for bandwidth which is appropriate for use with the orthogonal tapers we have already introduced.

3.9.1 Equivalent width of the approximate autocorrelation

Thomson (1977) shows that for f not near 0 or 1/2, and for a spectrum that is locally constant about f, the approximate correlation between $\hat{S}^{(d)}(f)$ and $\hat{S}^{(d)}(f+\eta)$, assuming that $\Delta t = 1$, is given by

$$R(\eta) = \left| \sum_{t=1}^{N} h_t^2 e^{-i2\pi\eta t} \right|^2 ,$$

which depends only on the data taper and not upon either f or $S(\cdot)$.

Since $R(\cdot)$ is a periodic function, its Fourier transform is a sequence $\{r_q\}$, given by

$$r_q \equiv \begin{cases} \sum_{t=1}^{N-|q|} h_t^2 h_{t+|q|}^2 & |q| \leq N-1 \\ 0 & |q| \geq N. \end{cases} \tag{3.23}$$

Our aim is to determine the 'cut-off' point of this approximate autocorrelation, that is, the point at which $R(\eta)$ reaches zero. One measure of this point is the equivalent width of $R(\eta)$, namely,

$$\text{width}_e\{R(\cdot)\} = \frac{\int_{-1/2}^{1/2} R(\eta)\, d\eta}{R(0)} = \int_{-1/2}^{1/2} R(\eta)\, d\eta.$$

This is a measure of the width of the rectangle whose height is equal to $R(0) = 1$, and whose area is the same as that under $R(\cdot)$. As $\{r_q\}$ and $R(\cdot)$ are a Fourier transform pair, we have

$$r_q = \int_{-1/2}^{1/2} R(\eta) e^{i2\pi\eta q}\, d\eta \ \text{ hence } \ r_0 = \int_{-1/2}^{1/2} R(\eta)\, d\eta.$$

From Equation 3.23, we have

$$\text{width}_e\{R(\cdot)\} = \int_{-1/2}^{1/2} R(\eta)\, d\eta = r_0 = \sum_{t=1}^{N} h_t^4. \tag{3.24}$$

Taper order	$\text{width}_e\{R(\cdot)\}$
0	1.9574
1	1.5314
2	1.4554
3	1.8512

Table 3.2: Values of $\text{width}_e\{R(\cdot)\}$ for dpss taper, $NW = 2$, $N = 32$

The values of $\text{width}_e\{R(\cdot)\}$ obtained for the dpss's with $NW = 2$, and $N = 32$, are given in Table 3.2. These values do not behave as one would expect. As we have already seen, as the order of the tapers increase, the grid of frequencies at which the direct spectral estimators are uncorrelated become more sparse. This is due to the fact that the difference in height between the main lobe of the spectral window and the first side lobe decreases as the order of dpss increases, as can be seen in Figure 3.4. However, $\text{width}_e\{R(\cdot)\}$ decreases before increasing as the order of taper goes up.

Because of the inadequacy of $\text{width}_e\{R(\cdot)\}$, an alternative bandwidth measure was developed as now discussed (see also Walden, McCoy & Percival, (1995)).

3.9.2 Autocorrelation width of the spectral window

In the case of a conventional lag window spectral estimator $\hat{S}^{(lw)}(\cdot)$,

$$E\{\hat{S}^{(lw)}(f)\} \approx \int_{-1/2}^{1/2} L(f - f')S(f')\, df',$$

where the smoothing window $L(\cdot)$ is the Fourier transform of the lag window. Jenkins (1961) defined the smoothing window bandwidth B_L as the equivalent width of the autocorrelation of $L(\cdot)$ – simply known as the autocorrelation width, which we will denote width_a.

The lag window spectral estimates $\hat{S}^{(lw)}(f)$ and $\hat{S}^{(lw)}(f+\eta)$ will be approximately uncorrelated when η exceed the smoothing window bandwidth. Hence we can utilise the smoothing window bandwidth as a measure of the cut-off point of the correlation of $\hat{S}^{(lw)}(f)$ and $\hat{S}^{(lw)}(f+\eta)$. The equivalent width of the real-function $L(\cdot)$ is given by

$$\begin{aligned}
\text{width}_a\{L(\cdot)\} &\equiv \text{width}_e\{L \star L(\cdot)\} \\
&= \frac{\int_{-1/2}^{1/2} L \star L(f)\, df}{L \star L(0)} \\
&= \frac{\left[\int_{-1/2}^{1/2} L(f)\, df\right]^2}{\int_{-1/2}^{1/2} L^2(f)\, df} \\
&= \left\{\int_{-1/2}^{1/2} L^2(f)\, df\right\}^{-1},
\end{aligned}$$

as the numerator is unity for all commonly used smoothing windows, and the *deterministic* cross-correlation denoted by $\star$ between $g(\cdot)$ and $h(\cdot)$ is defined as

$$g \star h(t) = \int_{-1/2}^{1/2} g(u)h(u+t)\, du.$$

This suggests that for direct spectral estimators we could define the spectral window bandwidth as the equivalent width of the autocorrelation of $\mathcal{H}(\cdot)$:

$$\begin{aligned}
\text{width}_a\{\mathcal{H}(\cdot)\} &= \frac{\left[\int_{-1/2}^{1/2} \mathcal{H}(f)\, df\right]^2}{\int_{-1/2}^{1/2} \mathcal{H}^2(f)\, df} \\
&= \left\{\int_{-1/2}^{1/2} \mathcal{H}^2(f)\, df\right\}^{-1}.
\end{aligned}$$

By Parseval's theorem

$$\int_{-1/2}^{1/2} \mathcal{H}^2(f)\, df = \sum_{\tau=-(N-1)}^{(N-1)} c_\tau^2,$$

where $\{c_\tau\} \longleftrightarrow \mathcal{H}(\cdot)$. But recall, for $\Delta t = 1$,

$$\mathcal{H}(f) = \left|\sum_{t=1}^{N} h_t e^{-i2\pi ft}\right|^2 = \sum_{\tau=-(N-1)}^{(N-1)} \left[\sum_{t=1}^{N-|\tau|} h_t h_{t+|\tau|}\right] e^{-i2\pi ft}.$$

Hence $c_\tau = \sum_{t=1}^{N-|\tau|} h_t h_{t+|\tau|}$. Thus

$$\begin{aligned}
\text{width}_a\{\mathcal{H}\} &= \left\{\int_{-1/2}^{1/2} \mathcal{H}^2(f)\, df\right\}^{-1} \\
&= \left\{\sum_{\tau=-(N-1)}^{(N-1)} c_\tau^2\right\}^{-1} \\
&= \left\{\sum_{\tau=-(N-1)}^{(N-1)} \left[\sum_{t=1}^{N-|\tau|} h_t h_{t+|\tau|}\right]^2\right\}^{-1}. \qquad (3.25)
\end{aligned}$$

Now width_a can be used in exactly the same way as B_L to measures the cut-off point of the correlation between $\hat{S}^{(d)}(f)$ and $\hat{S}^{(d)}(f+\eta)$. Notice also from Equation 3.25 that the autocorrelation of $\mathcal{H}$ is well-defined and can easily be calculated for any given taper.

3.9.3 Comparison of width_a and width_e

Figure 3.19 shows both width_a and width_e for the discrete prolate spheroidal sequences with $NW = 2, 3, 4$ (the shortest line corresponding to $NW = 2$, and the longest line corresponding to $NW = 4$) and $N = 32$. The measure $\text{width}_e\{R(\cdot)\}$ is dominated by the main lobe of $R(\cdot)$; however, $\text{width}_a\{\mathcal{H}(\cdot)\}$ increases with increasing taper order due to the prominent sidelobes of the autocorrelation of $\mathcal{H}(\cdot)$.

The nature of the discrete prolate spheroidal sequences means that high orders, especially those close to the Shannon number ($2NW$), tend to have a preponderance of energy outside the main lobe, which can lead to erratic behaviour in $\text{width}_a\{\mathcal{H}(\cdot)\}$ for higher order tapers. This must be balanced against the fact that the very behaviour of these higher order tapers would tend to be preclude them from any multitaper estimates, so the erratic behaviour of $\text{width}_a\{\mathcal{H}(\cdot)\}$ in this respect is unlikely be a practical problem.

Figure 3.20 shows both width_a and width_e for the discrete sinusoidal tapers for $N = 32$, and in this case $\text{width}_e\{R(\cdot)\}$ remains constant with increasing order, while $\text{width}_a\{\mathcal{H}(\cdot)\}$ increases with increasing order.

3.9.4 Cut-off of exact spectral correlation

In order to establish the accuracy of the two bandwidth measures outlined in Equations 3.24 and 3.25, we can compare them to the exact correlation between $\hat{S}^{(d)}(0.25)$ and $\hat{S}^{(d)}(0.25+\eta)$ for $-0.25 < \eta < 0.25$, for white noise. This is purely for illustrative purposes, as in practice we would have no need to taper a white noise process.

For white noise with variance σ^2, the exact covariance between direct spectral

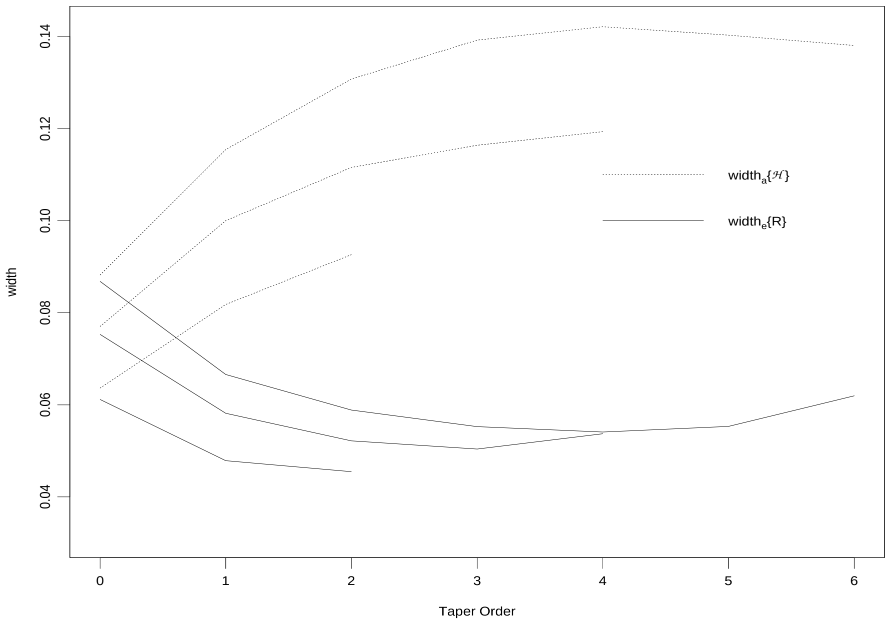


Figure 3.19: Comparison of equivalent widths for discrete prolate spheroidal sequences

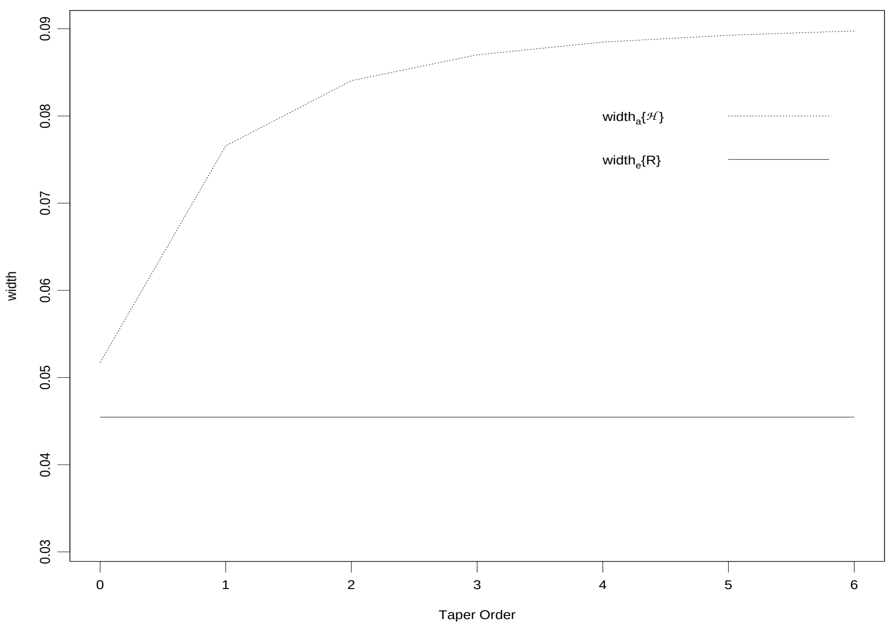


Figure 3.20: Comparison of equivalent widths for sinusoidal tapers

estimators is (see Section 3.8 for details)

$$\operatorname{cov}\{\hat{S}^{(d)}(f), \hat{S}^{(d)}(f+\eta)\} = \sigma^4 \left\{ \left| \sum_{t=1}^{N} h_t^2 e^{i2\pi\eta t} \right|^2 + \left| \sum_{t=1}^{N} h_t^2 e^{i2\pi(2ft+\eta t)} \right|^2 \right\},$$

from which the exact correlation follows. Figure 3.21 shows the exact correlation between $\hat{S}^{(d)}(0.25)$ and $\hat{S}^{(d)}(0.25+\eta)$ plotted as the solid line, using the discrete prolate spheroidal sequences with $NW = 2$ and $N = 32$, while the two vertical lines depict the two bandwidth measures. We can see from these plots that our preferred measure $\text{width}_a\{\mathcal{H}(\cdot)\}$ captures the sidelobes of the correlation which $\text{width}_e\{R(\cdot)\}$ misses, and thus provides a more sensible measure.

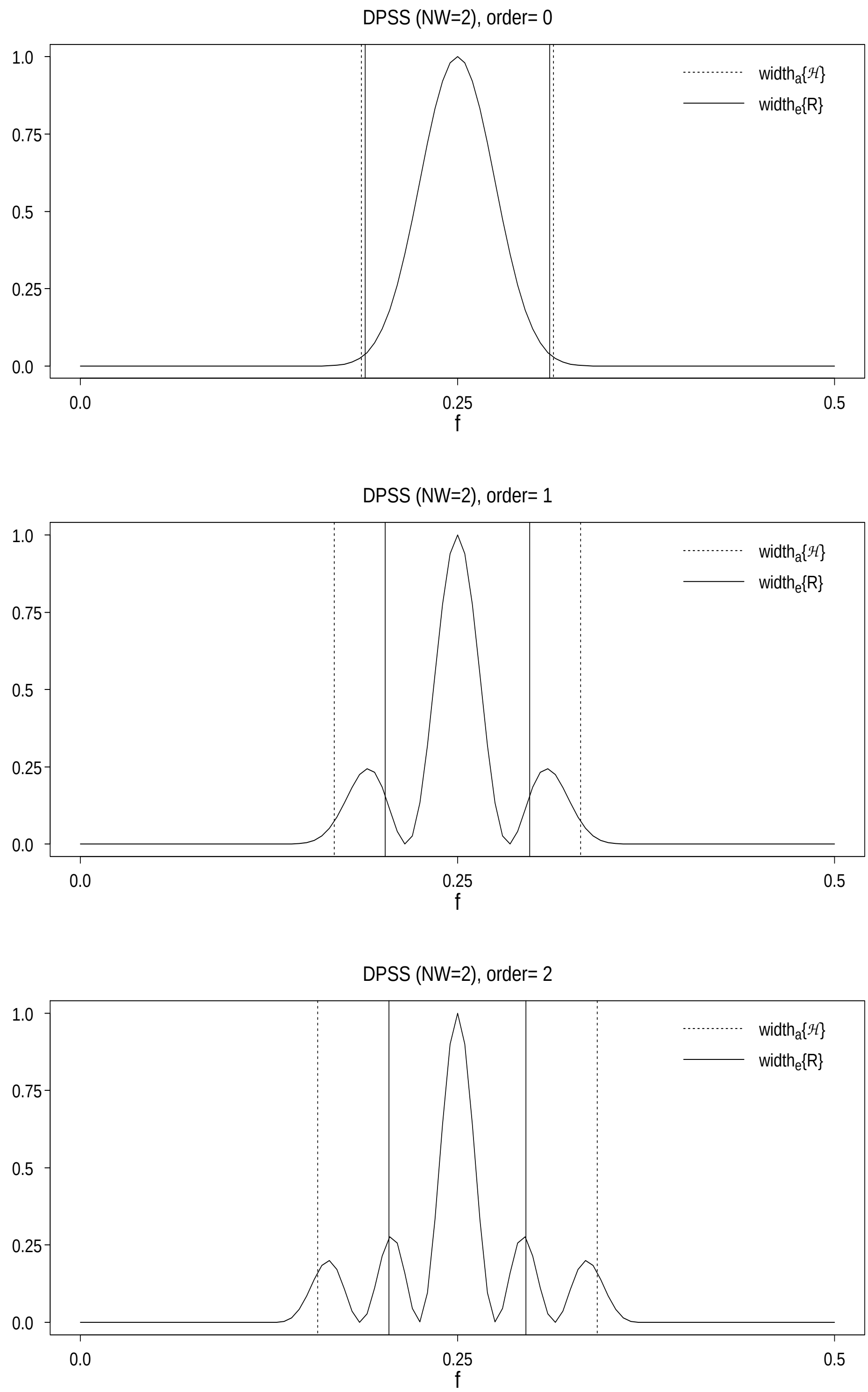


Figure 3.21: Comparison of width measures against actual correlations

Chapter 4

Multitaper Spectral Estimation

4.1 Introduction

Although we have seen that tapering is a means of producing an estimator with acceptable bias properties, we have also pointed out there is a price to pay in that the variance efficiency is reduced. There are alternative estimation procedures, such as prewhitening or Welch's overlapped segment averaging, which avoid this inherent loss of information.

Here we outline another approach to spectral estimation, first introduced by Thomson (1982), which involves the averaging of several direct spectral estimators formed using orthogonal data tapers. We shall refer to this method of spectral estimation as multitaper spectral estimation and, as an estimator, it has several properties in its favour. Multitapering is becoming an increasingly popular method of spectral analysis in cases where the true spectrum is rapidly varying.

Multitapering is an automatic procedure which can be applied to any data set without, for example, the design of an individual prewhitening filter for each data set.

The performance of multitaper spectral analysis was compared to another spectral analysis technique – Welch's overlapped segment averaging by Bronez (1992). He compared the methods using three criteria; leakage, variance and resolution, and found that evaluating one of these criteria while the remaining two were fixed, resulted in multitapering always producing a preferred measure for the varying criteria. Bronez (1988) has also extended the method of multitaper spectral estimation to analyzing irregularly sampled multidimensional processes.

Multitapering has already proved useful in applications such as the relationship between atmospheric carbon dioxide and global temperature (Kuo *et al* 1990) and investigating the spectral properties of geophysical data (Walden (1994)).

The bias of multitaper estimators can be broken down into local or smoothing bias, that is the bias within the smoothing interval, and broadband bias, the bias outside this interval.

One of the problems of using a single taper is that it gives lower weight to the end points of the series, whilst this serves the purpose of reducing bias, the information contained in the series is not being used in full. As the orthogonal tapers used in multitapering highlight different parts of the data, there is some justification in claiming that multitapering restores this lost information, without losing the good bias properties that tapering provides.

A full discussion of the method of multitaper spectral estimation, and two alternative justifications for the method; one using the theory of quadratic spectral estimators, and the other following more closely Thomson's original derivation, can be found in Percival & Walden (1993).

New work in this chapter include developing the exact form of the variance of a multitaper estimator, and extending the bandwidth measure we developed for direct spectral estimators to the multitaper case. We also show that the distributional assumptions for multitaper spectral estimators break down at low frequencies for certain long memory processes.

4.2 A simple multitaper estimator

Suppose we have a time series $X_1, X_2, \ldots, X_N$ which is a realization of the zero mean stationary process $\{X_t\}$, with variance σ^2 and sdf $S(f)$. Assume that the sampling interval is given by Δt, giving a Nyquist frequency of $f_{(N)} \equiv 1/2\Delta t$. In its simplest formulation, the multitaper estimator is simply the average of K

direct spectral estimators:

$$\begin{aligned} \hat{S}^{(mt)}(f) &\equiv \frac{1}{K}\sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f), \\ \text{where} \qquad \hat{S}_k^{(mt)}(f) &\equiv \Delta t \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi f t \Delta t} \right|^2, \end{aligned}$$

and $\{h_{t,k}\}$ is the data taper for the kth direct spectral estimator $\hat{S}_k^{(mt)}(f)$. For each data taper there is an associated spectral window $\mathcal{H}_k(f)$. It follows that

$$\mathrm{E}\{\hat{S}^{(mt)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \overline{\mathcal{H}}(f - f')S(f')\,df' \text{ with } \overline{\mathcal{H}}(f) \equiv \frac{1}{K}\sum_{k=0}^{K-1} \mathcal{H}_k(f),$$

where $\overline{\mathcal{H}}(f)$ is the overall spectral window for the multitaper estimator, $\hat{S}^{(mt)}(f)$. To ensure that the resulting individual spectra are as uncorrelated as possible the data tapers should be orthogonal, and of course the individual tapers should exhibit good leakage properties.

4.2.1 Multitaper estimators using the discrete prolate spheroidal sequences

An obvious choice for a set of orthogonal tapers is the discrete prolate spheroidal sequences with parameter W and orders $k = 0$ to $K - 1$, where K is chosen to be less than the Shannon number $2NW\Delta t$, which we introduced in Section 3.4.1.

Here, W should be chosen such that $0 < W < f_{(N)}$, usually W is defined implicitly by choosing the parameter $NW\Delta t$ to be some small value, say $NW\Delta t = 2, 3$ or 4. We can then form the individual direct spectral estimates and average together K of these estimates, where K is determined from examination of the individual estimates, but is strictly less than the Shannon number to ensure good leakage properties.

The mean spectral windows $\overline{\mathcal{H}}(\cdot)$ when using the discrete prolate spheroidal sequences with $NW\Delta t = 4$, $\Delta t = 1$, $N = 32$, and an increasing number of tapers is shown in Figure 4.1, where the vertical line shows the half bandwidth W.

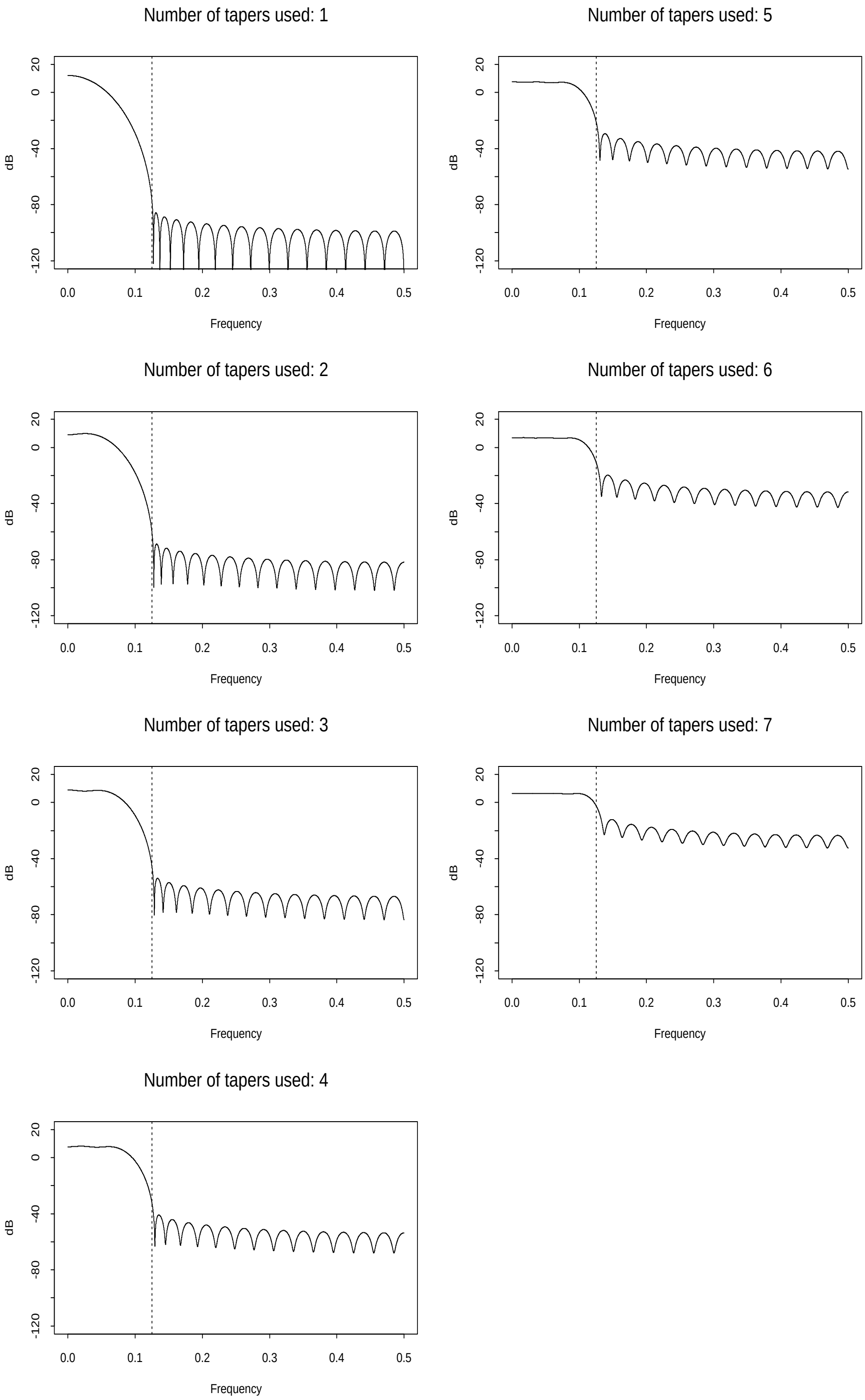


Figure 4.1: Mean spectral window for multitapering using dpss's, $NW = 4$, $N = 32$

4.2.2 Multitaper estimators using the discrete sinusoidal tapers

An alternative choice for a set of orthogonal tapers is the discrete sinusoidal tapers introduced in Section 3.4.2. Here, all the tapers have the same bandwidth, but they are centred at different frequencies. Increasing the number of tapers used in any multitaper spectral estimator has the effect of increasing the bandwidth, until the desired amount of local smoothing is achieved. The mean spectral windows $\overline{\mathcal{H}}(\cdot)$ when using the discrete sinusoidal tapers with $N = 32$, $\Delta t = 1$ and an increasing number of tapers is shown in Figure 4.2. The vertical line shows the value of $(k+1)/2(N+1)$, where k is the number of tapers used (this defines the half bandwidth of the mean spectral window - from Reidel & Sidorenko (1994)).

4.3 Statistical properties of multitaper spectral estimators

From the known approximate distribution of the individual direct spectral estimators, it immediately follows that, conditional on $\hat{S}_k^{(mt)}(f)$ being uncorrelated,

$$\hat{S}^{(mt)}(f) \stackrel{d}{=} \begin{cases} [S(f)/2K]\,\chi^2_{2K} & 0 < f < f_{(N)} \\ [S(f)/K]\,\chi^2_K & f = 0 \text{ or } f_{(N)}. \end{cases}$$

This implies that

$$\mathrm{var}\{\hat{S}^{(mt)}(f)\} = \begin{cases} S^2(f)/K & 0 < f < f_{(N)} \\ 2S^2(f)/K & f = 0 \text{ or } f_{(N)}, \end{cases}$$

thus reducing the variance of the estimator by a factor of K over the individual direct spectral estimators. If $\{X_t\}$ is Gaussian, then the χ^2 distribution is exact, otherwise it is an approximation which follows from the central limit theorem.

In Chapter 5 we will be employing the use of multitaper estimators to obtain parameter estimates of fractionally differenced processes, using estimates of the spectrum on a decibel ($10\log_{10}$) scale, in this case letting

$$Z = 10\log_{10}(\hat{S}^{(mt)}(f)),$$

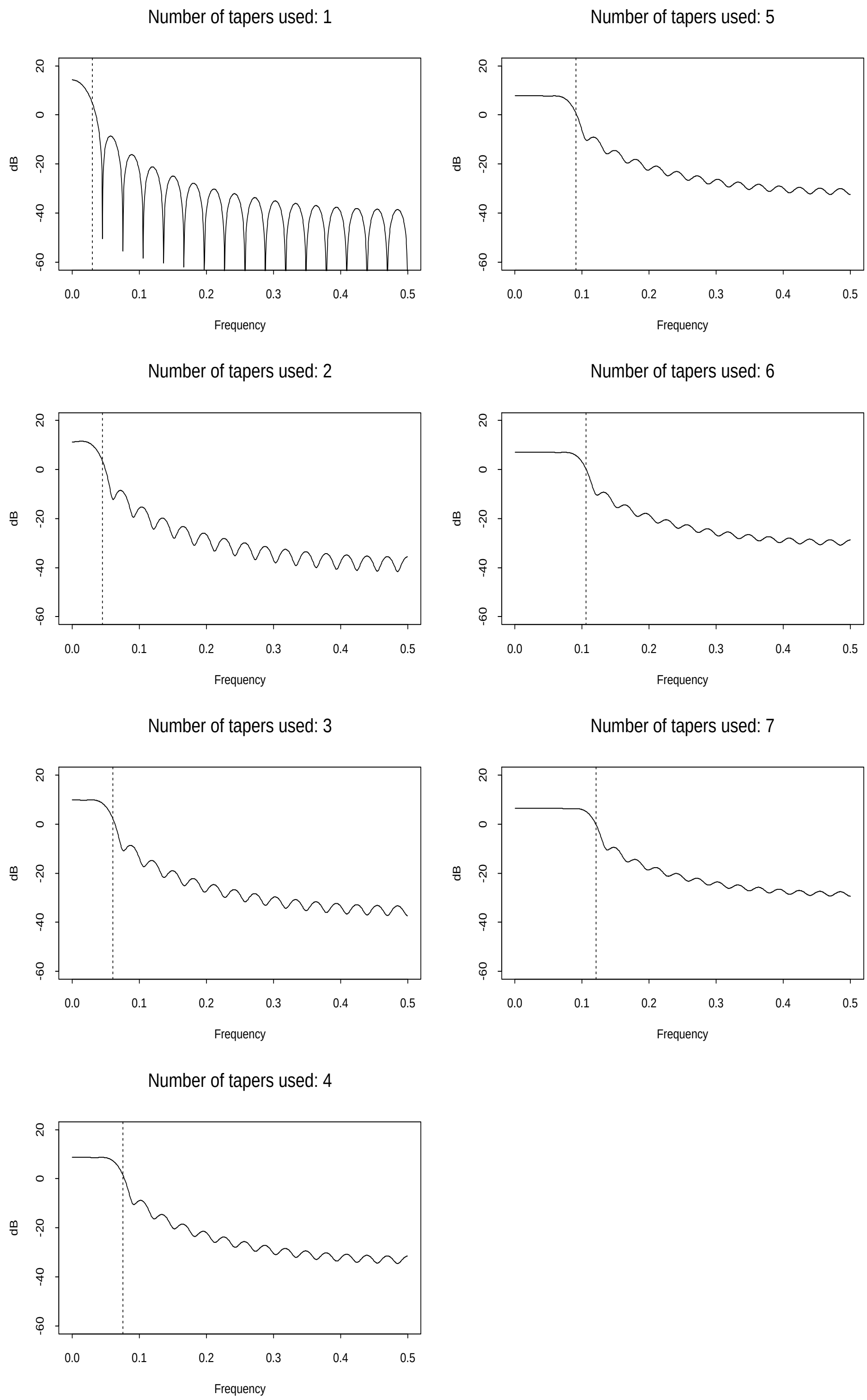


Figure 4.2: Mean spectral window for multitapering using the discrete sinusoidal tapers

then, the probability density function of Z is given by

$$f_z(Z) = \frac{10^{\frac{z}{10}}}{10} \log(10) f_y\left(10^{\frac{z}{10}}\right),$$

where $f_y(y)$ is the probability density function corresponding to a χ^2_{2K} distribution. We will denote the probability density function of Z as $f_z(Z) = 10 \log_{10} \chi^2_{2K}$.

This density is shown in Figure 4.3 for increasing values of K. We see that the effect of the increase in degrees of freedom results in a more symmetrical density function, which also becomes more Gaussian (Bartlett & Kendall, 1946).

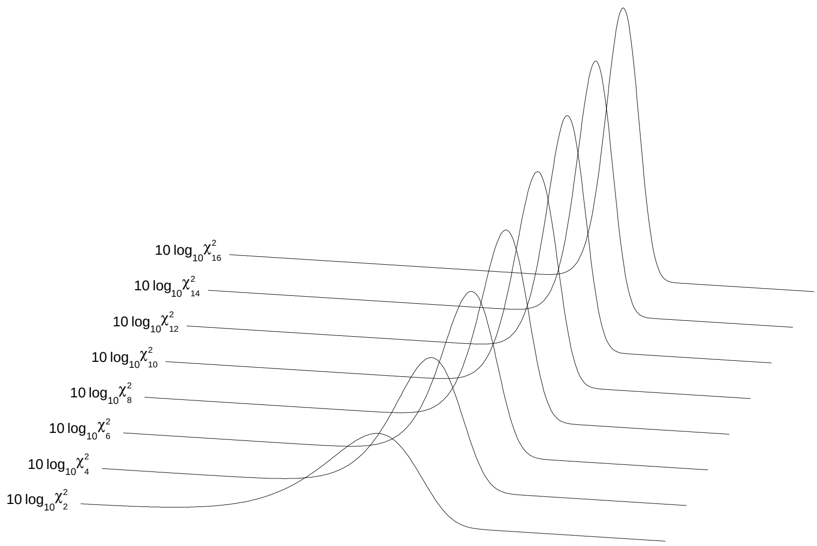


Figure 4.3: Probability density function for $10 \log_{10} \chi^2$ distributions

Figure 4.4 shows the multitaper estimates of the data shown in Figure 2.4, formed using the discrete prolate spheroidal sequences. Here, we have used the dpss's as tapers with $NW\Delta t = 4, K = 5$, and $\Delta t = 1$. It is clear from the plots that the variance of the estimate is lower than that of the periodogram.

Figure 4.5 shows the multitaper estimates of the data shown in Figure 2.4, formed using the discrete sinusoidal tapers with $K = 5$. Again we see a reduction in the variance of the estimate.

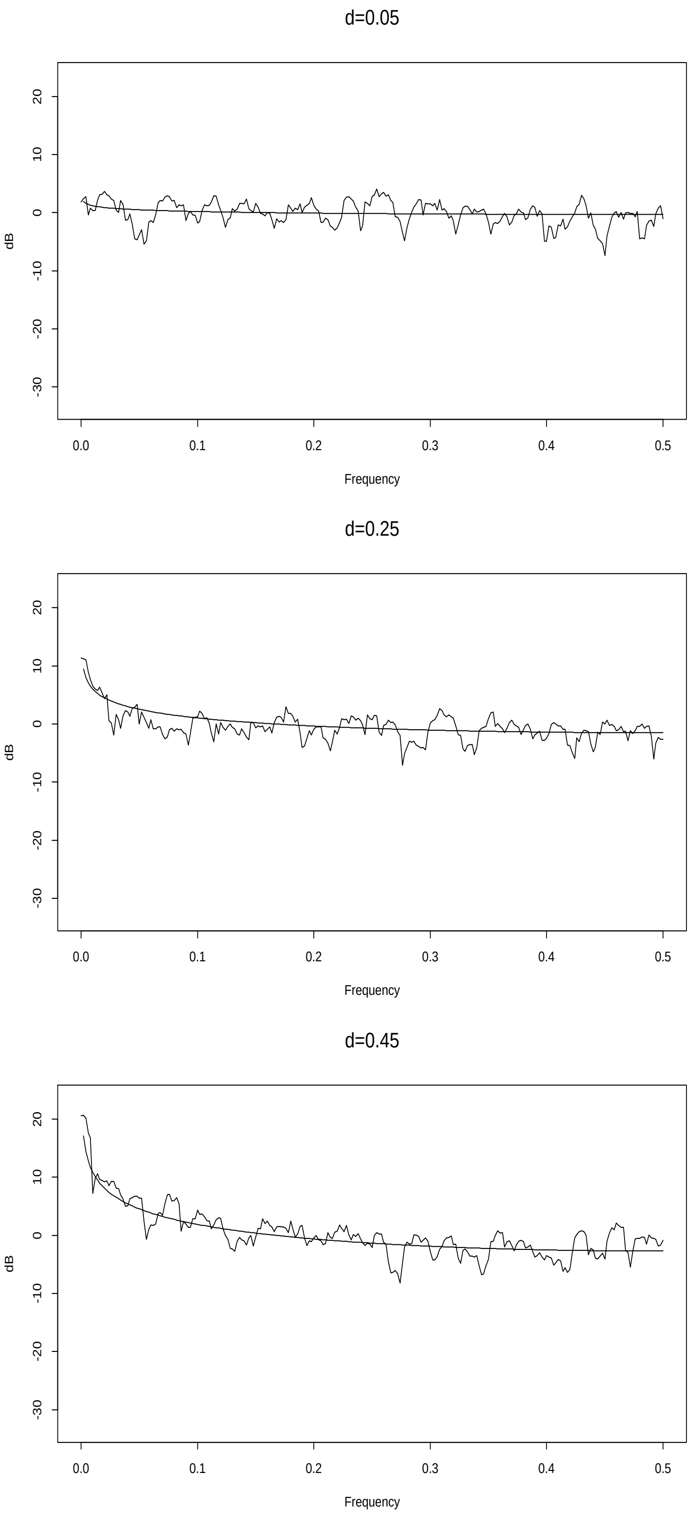


Figure 4.4: Dpss multitaper estimates of fractionally differenced processes

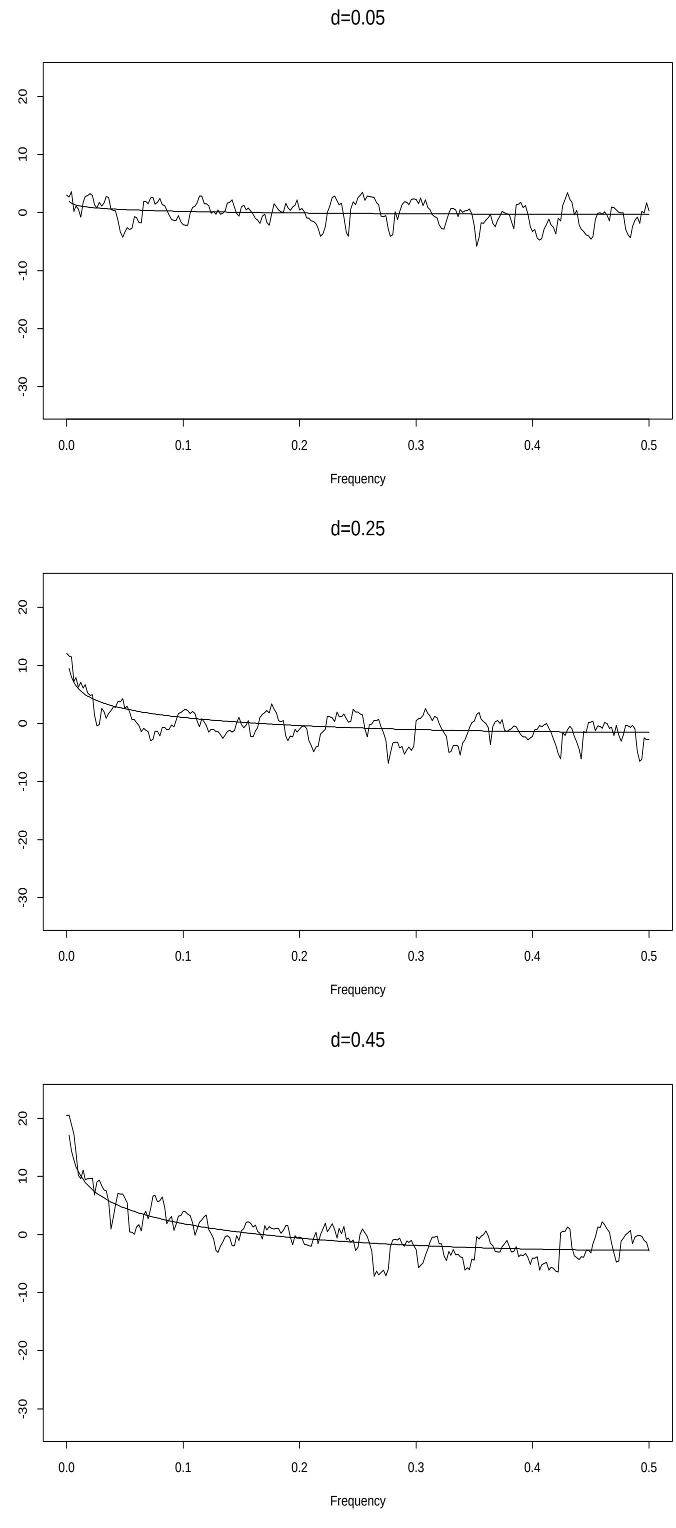


Figure 4.5: Sine multitaper estimates of fractionally differenced processes

4.3.1 Variance of multitaper estimators

We have assumed that choosing a set of orthogonal estimators ensures that the K direct spectral estimators are uncorrelated, but how approximate is this assumption? In Walden, McCoy & Percival (1994), we developed the exact variance of a multitaper spectral estimator, which we shall now outline.

We know that

$$\begin{aligned}\operatorname{var}\{\hat{S}^{(mt)}(f)\} &= \operatorname{var}\left\{\frac{1}{K}\sum_{k=0}^{K-1}\hat{S}_k^{(mt)}(f)\right\} \\ &= \frac{1}{K^2}\sum_{j=0}^{K-1}\sum_{k=0}^{K-1}\operatorname{cov}\{\hat{S}_j^{(mt)}(f),\hat{S}_k^{(mt)}(f)\}. \qquad (4.1)\end{aligned}$$

Assuming that $\Delta t = 1$ and letting

$$d_j(f) = \sum_{t=1}^{N} h_{t,j}X_t e^{-i2\pi ft} \qquad \text{and} \qquad d_k(f) = \sum_{u=1}^{N} h_{u,k}X_t e^{-i2\pi fu},$$

then

$$\hat{S}_j^{(mt)}(f) = |d_j(f)|^2 \qquad \text{and} \qquad \hat{S}_k^{(mt)}(f) = |d_k(f)|^2.$$

It follows from a version of the Isserlis theorem (see Section 3.8), that

$$\begin{aligned}\operatorname{cov}\{\hat{S}_j^{(mt)}(f),\hat{S}_k^{(mt)}(f)\} &= \operatorname{cov}\{|d_j(f)|^2,|d_k(f)|^2\} \\ &= \operatorname{cov}\{d_j(f)d_j^*(f),d_k(f)d_k^*(f)\} \\ &= E\{d_j(f)d_k(f)\}E\{d_j^*(f)d_k^*(f)\} \\ &\quad + E\{d_j(f)d_k^*(f)\}E\{d_j^*(f)d_k(f)\} \\ &= |E\{d_j(f)d_k(f)\}|^2 + |E\{d_j^*(f)d_k(f)\}|^2,\end{aligned}$$

which, after some algebra gives us the following form of the covariance between two direct spectral estimators

$$\begin{aligned}\operatorname{cov}\{\hat{S}_j^{(mt)}(f),\hat{S}_k^{(mt)}(f)\} &= \left|\sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u-t)}\right|^2 \\ &\quad + \left|\sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u+t)}\right|^2. \qquad (4.2)\end{aligned}$$

Hence, from Equation 4.1

$$\begin{aligned}\operatorname{var}\{\hat{S}^{(mt)}(f)\} &= \frac{1}{K^2}\left\{\sum_{j=0}^{K-1}\sum_{k=0}^{K-1}\left|\sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u-t)}\right|^2\right\} \\ &\quad + \left\{\sum_{j=0}^{K-1}\sum_{k=0}^{K-1}\left|\sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u+t)}\right|^2\right\}. \qquad (4.3)\end{aligned}$$

1. Use a forward FFT to give

$$Y_{t,k}(f_l) = \frac{1}{N}\sum_{u=1}^{N}(h_{u,k}s_{t-u})e^{-i2\pi u f_l},$$

where $f_l = l/N$.

2. For the first term of Equation 4.3 apply an inverse FFT to $\{h_{t,j}Y_{t,k}(f_l)\}$ and multiply by N to obtain

$$N\left\{\sum_{t=1}^{N} h_{t,j}Y_{t,k}(f_l)e^{i2\pi t f_l}\right\} = \sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u-t)}.$$

3. For the second term of Equation 4.3 apply a forward FFT to $\{h_{t,j}Y_{t,k}(f_l)\}$ and multiply by N^2 to obtain

$$N^2\left\{\frac{1}{N}\sum_{t=1}^{N} h_{t,j}Y_{t,k}(f_l)e^{-i2\pi t f_l}\right\} = \sum_{t=1}^{N}\sum_{u=1}^{N} h_{t,j}h_{u,k}s_{t-u}e^{-i2\pi f(u+t)}.$$

Table 4.1: Algorithm for calculating the variance of multitaper spectral estimators

This expression gives us a means for assessing how correlated the actual individual estimators are at specific frequencies given the autocovariance sequence (acvs). So, for example, we are in a position to examine the covariance structure of the estimators for different processes such as white noise or fractionally differenced processes. If we were using dpss's as the tapers, we would expect the covariances to be negligible for $K < 2NW$. A method for calculating the expression given in Equation 4.3 using fast Fourier transforms is given in Table 4.1.

4.3.2 Variance of multitaper estimators for white noise

Although it is not necessary to use multitapering in the case of white noise (as it has no dynamic range), it serves to illustrate the use of the expression for the variance given in the previous section. For simplicity, we take white noise with a variance of unity, so that the acvs is zero everywhere except at lag zero where it is unity. Using the exact form of the acvs we can calculate the variance of the multitaper spectral estimator using Equations 4.1 and 4.2 (see also Walden, McCoy & Percival (1994)). These variances are shown in Figure 4.6, with $NW = 4$ and $N = 32$. The figures show the theoretical variance of the multitaper estimators:

$$\mathrm{var}\{\hat{S}^{(mt)}(f)\} = \begin{cases} S^2(f)/K & 0 < f < f_{(N)} \\ 2S^2(f)/K & f = 0 \text{ or } f_{(N)}, \end{cases}$$

as well as the true variance made up by either including or excluding the covariances between the individual direct spectral estimators.

Away from zero frequency, the individual direct spectral estimators are uncorrelated below $K = 2NW$ (marked by the vertical line) at all frequencies, and the variance matches the theoretical variance. At frequencies close to zero, typically less than the design bandwidth W, we can see that correlations exist between the individual direct spectral estimators.

Figure 4.7 shows the variance of multitaper estimators of white noise when the discrete sinusoidal tapers are used. Here we see a similar pattern as for the dpss's, though at the low frequencies the variance follows the true variance for a higher number of tapers. This can be explained by examining the mean spectral windows for the multitaper estimators as shown in Figures 4.1 and 4.2, which show that the build up of the mean spectral window for the sinusoidal tapers occurs at a slower rate than that of the dpss's.

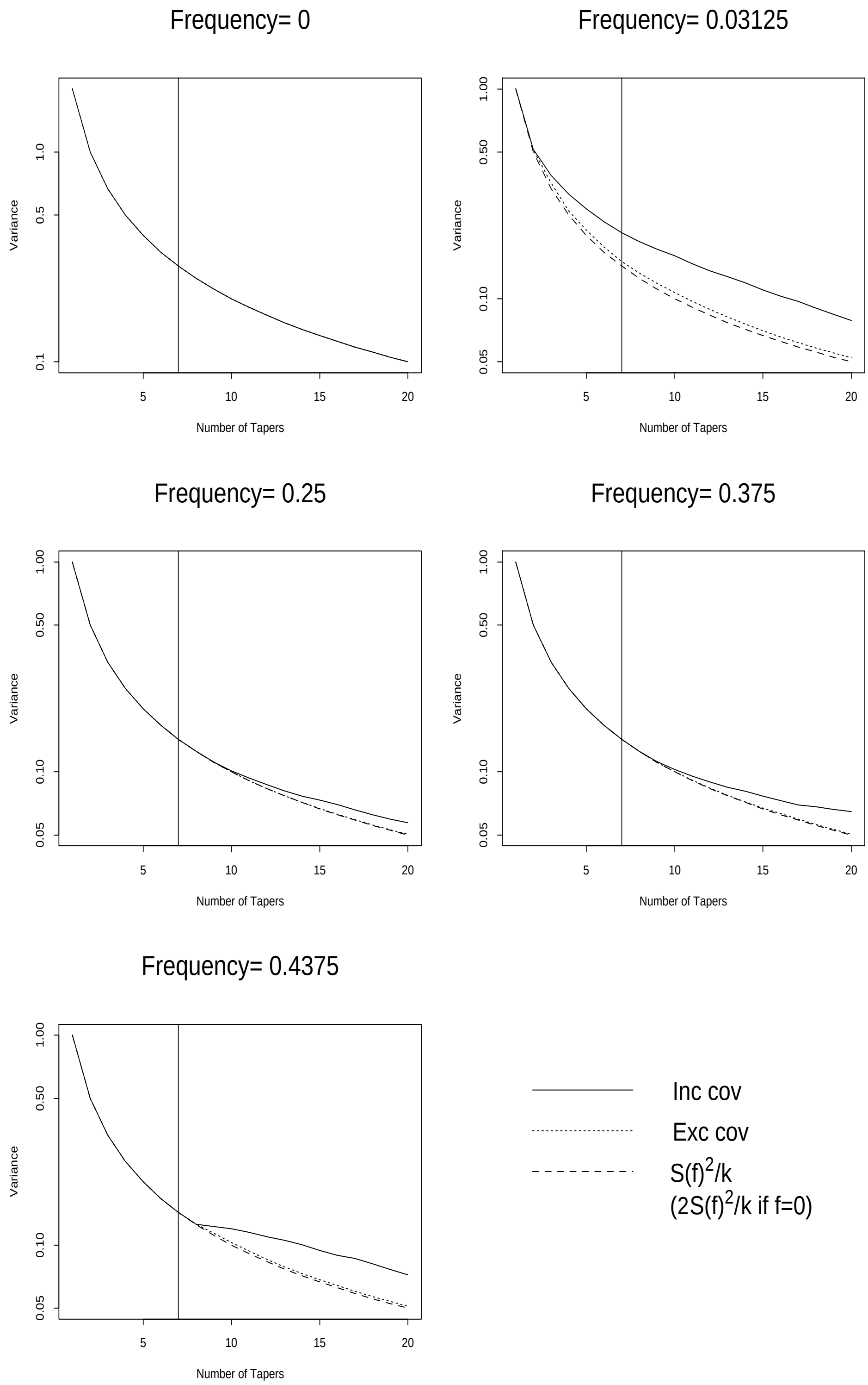


Figure 4.6: Variance of DPSS multitaper estimators of white noise

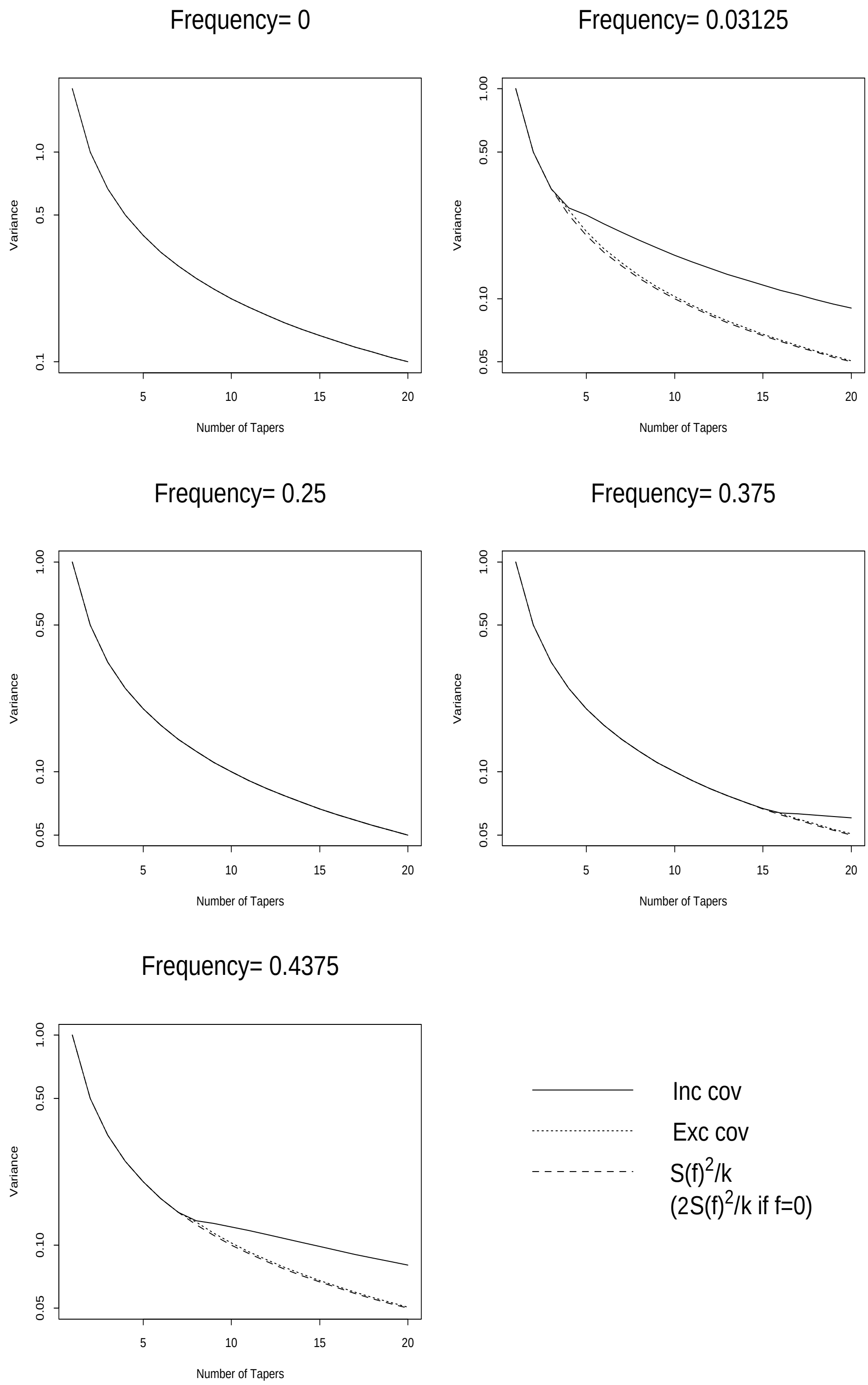


Figure 4.7: Variance of sinusoidal multitaper estimators of white noise

4.3.3 Variance of multitaper estimators for fractionally differenced processes

For a fractionally differenced process (fdp), the acvs is given by

$$s_\tau = \frac{\sigma_\epsilon^2 (-1)^\tau (-2d)!}{(\tau - d)!(-\tau - d)!},$$

we can then substitute this expression into Equations 4.1 and 4.2 to assess the independence of the multitaper estimators of fdp's. Due to the nature of fdp's, as d increases so does the steepness of the ascent of the spectrum in the region approaching zero frequency.

Figures 4.8 to 4.10 show the variances of the multitaper estimators of fractionally differenced processes with varying values of d, calculated both including and excluding the covariances between the individual direct spectral estimators; and using the discrete prolate spheroidal sequences as tapers with $NW = 4$, $N = 32$, and $\sigma_\epsilon^2 = 1$. For $d = 0.05$, the variance of the estimator behaves in a similar way to white noise with the covariances remaining negligible for $K < 2NW$, provided the frequency is sufficiently far from zero, where there are bias problems due to the nature of the sdf. For $K \geq 2NW$ the variances rise dramatically. As d increases the variance of the estimator become more and more inflated at frequencies close to zero.

The theoretical variance $S^2(f)/K$, or $2S^2(f)/K$ if the frequency is either 0 or 0.5, is also shown in the graphs, and we can see that as d increases the variance of the estimator near zero frequency is much higher than we would expect.

Figures 4.11 to 4.13 show the variances of the multitaper estimators of fractionally differenced processes with varying values of d, using the discrete sinusoidal tapers. As was the case for white noise, these variances appear to be closer to the theoretical variance at the lower frequencies when using a low number of tapers than estimators calculated using the discrete prolate spheroidal sequences. This is again due to the slower build up in the mean spectral window for the sinusoidal tapers.

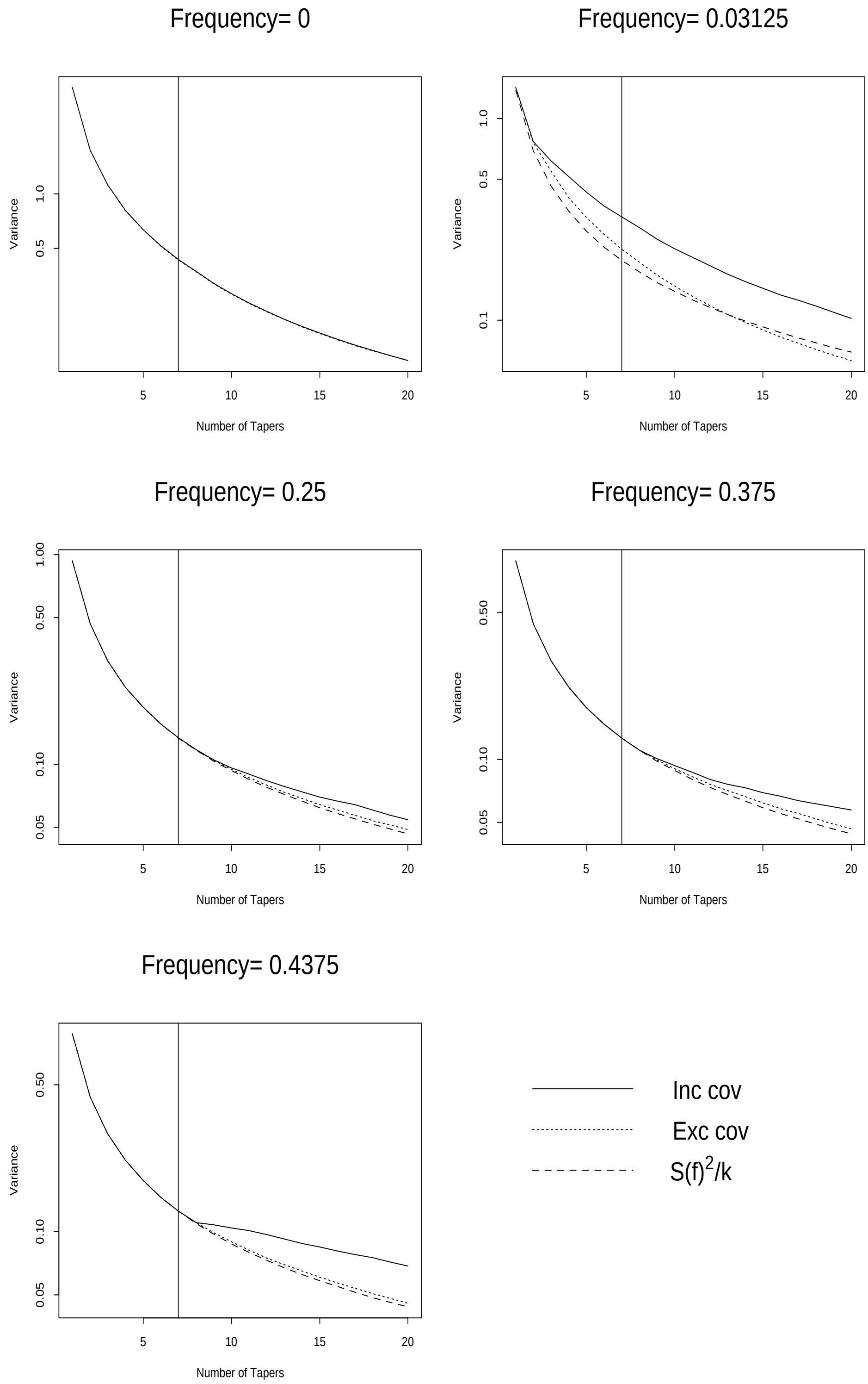


Figure 4.8: Variance of DPSS multitaper estimators of an fdp(0.05)

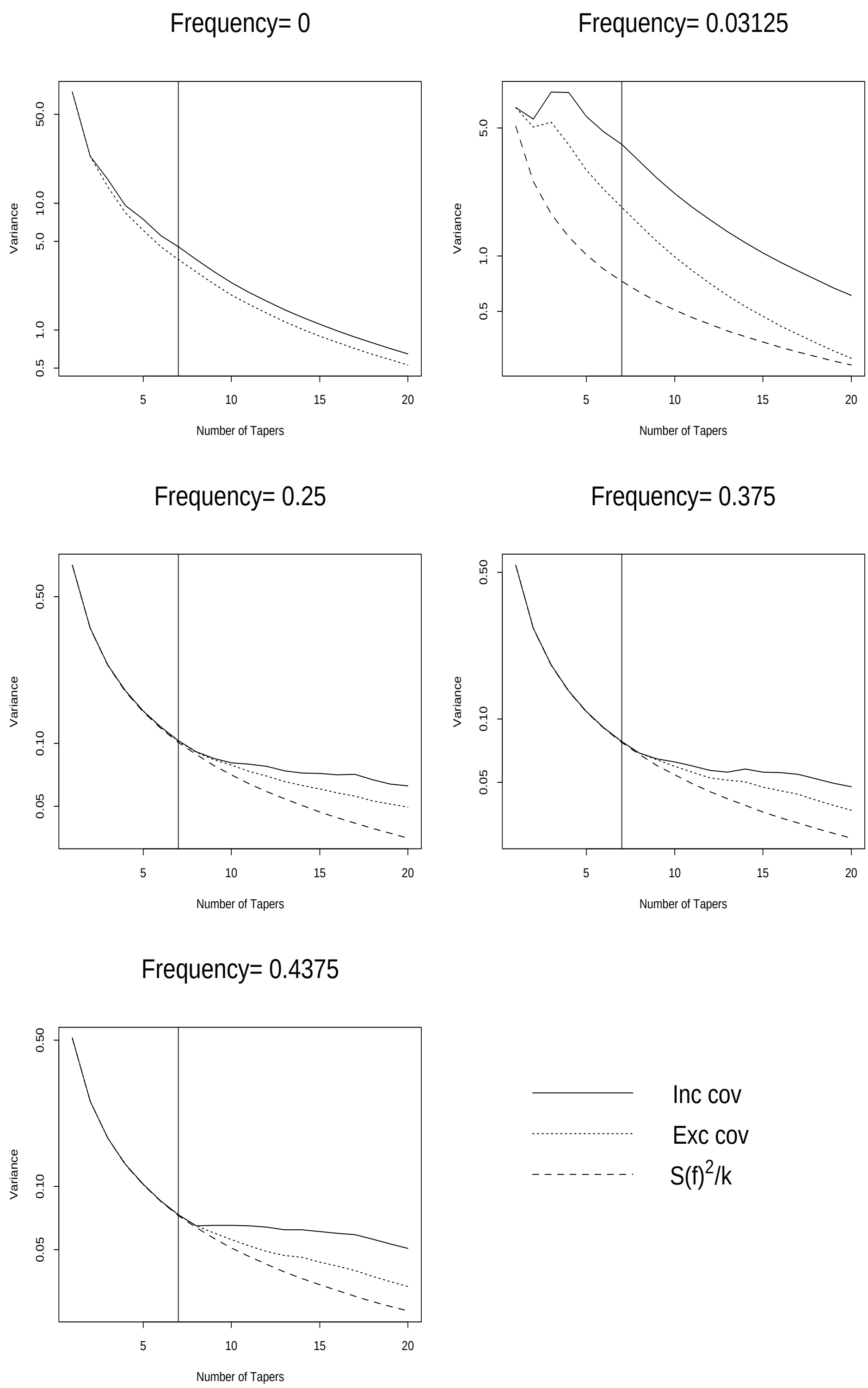


Figure 4.9: Variance of DPSS multitaper estimators of an fdp(0.25)

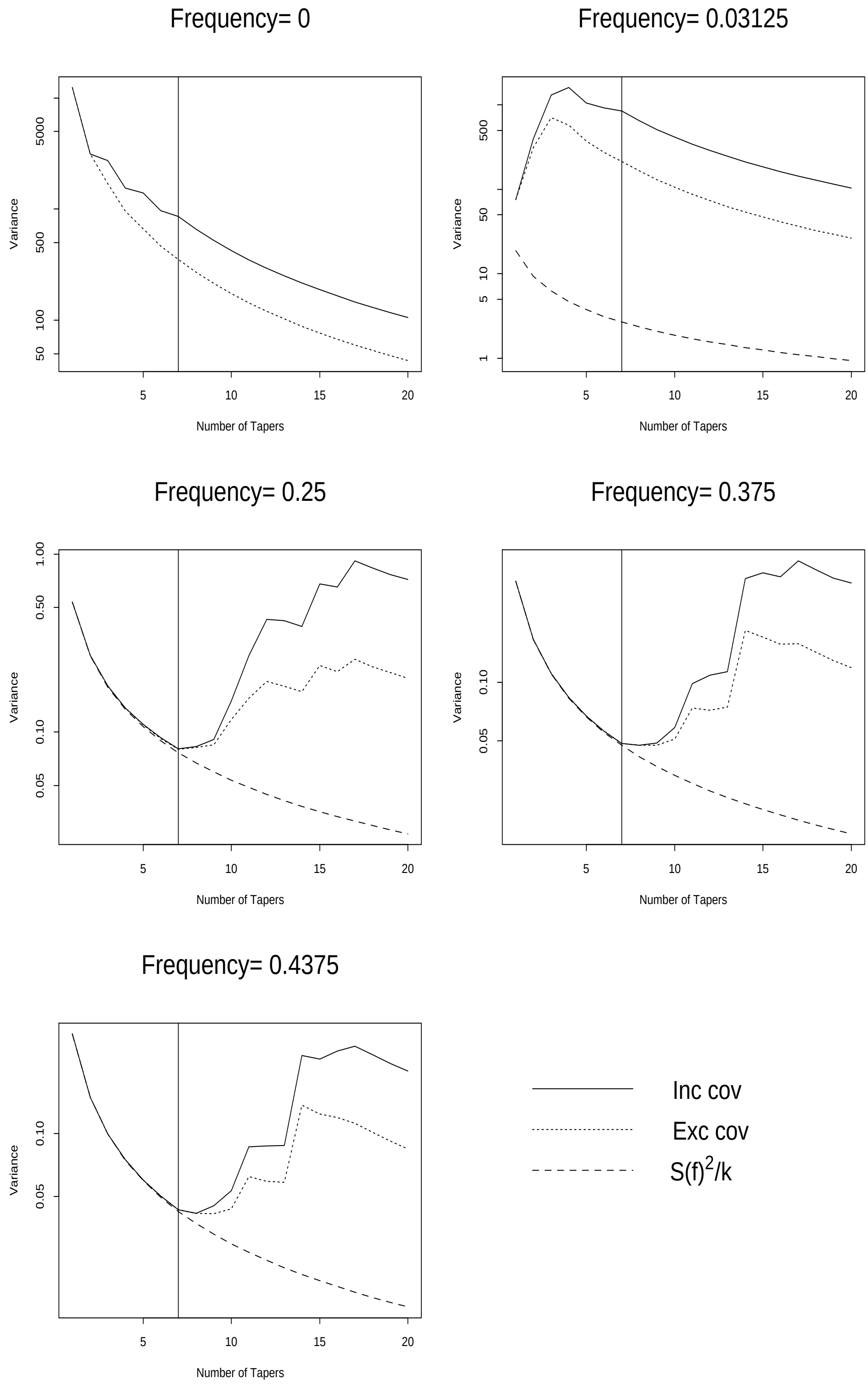


Figure 4.10: Variance of DPSS multitaper estimators of an fdp(0.45)

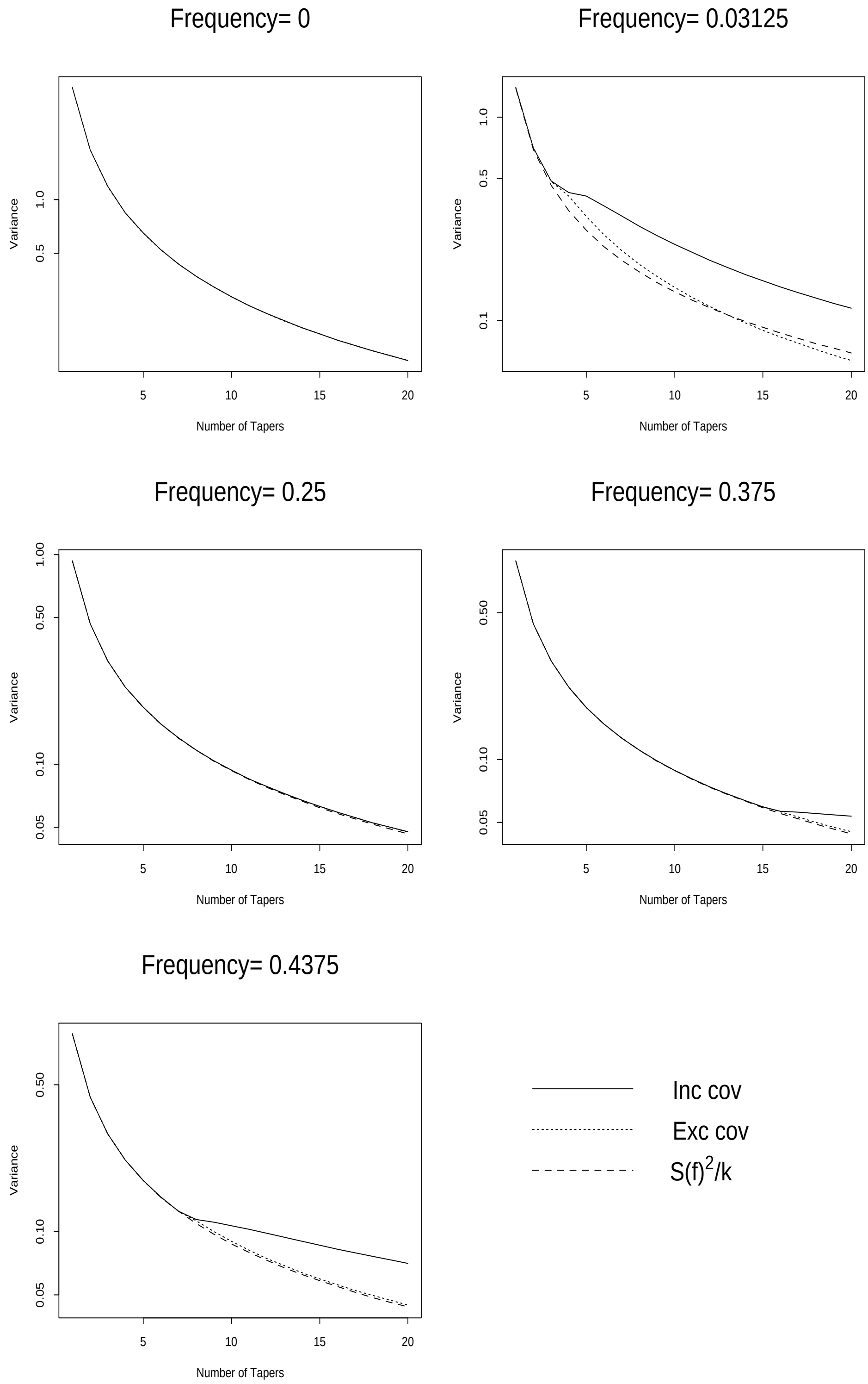


Figure 4.11: Variance of sinusoidal multitaper estimators of an fdp(0.05)

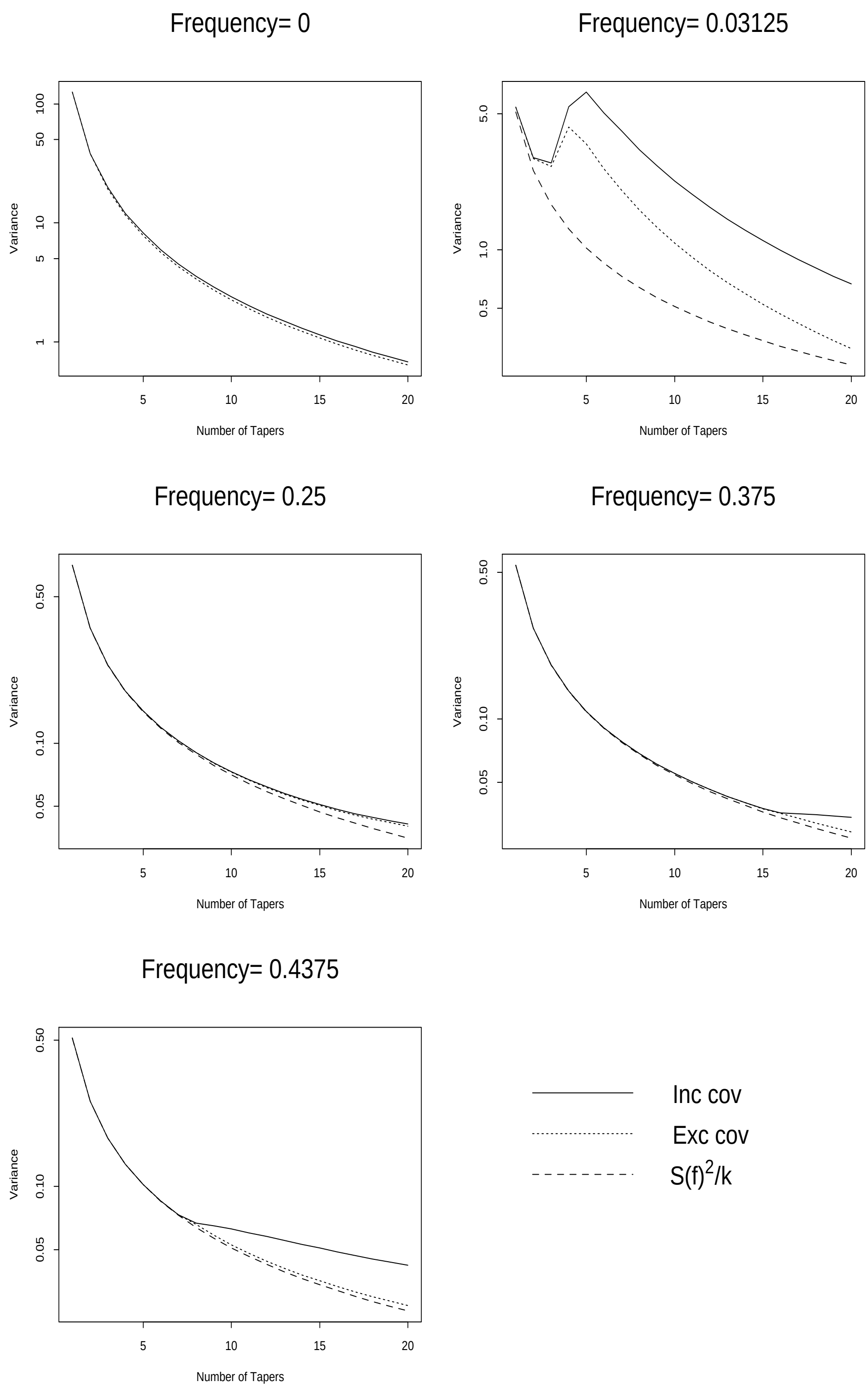


Figure 4.12: Variance of sinusoidal multitaper estimators of an fdp(0.25)

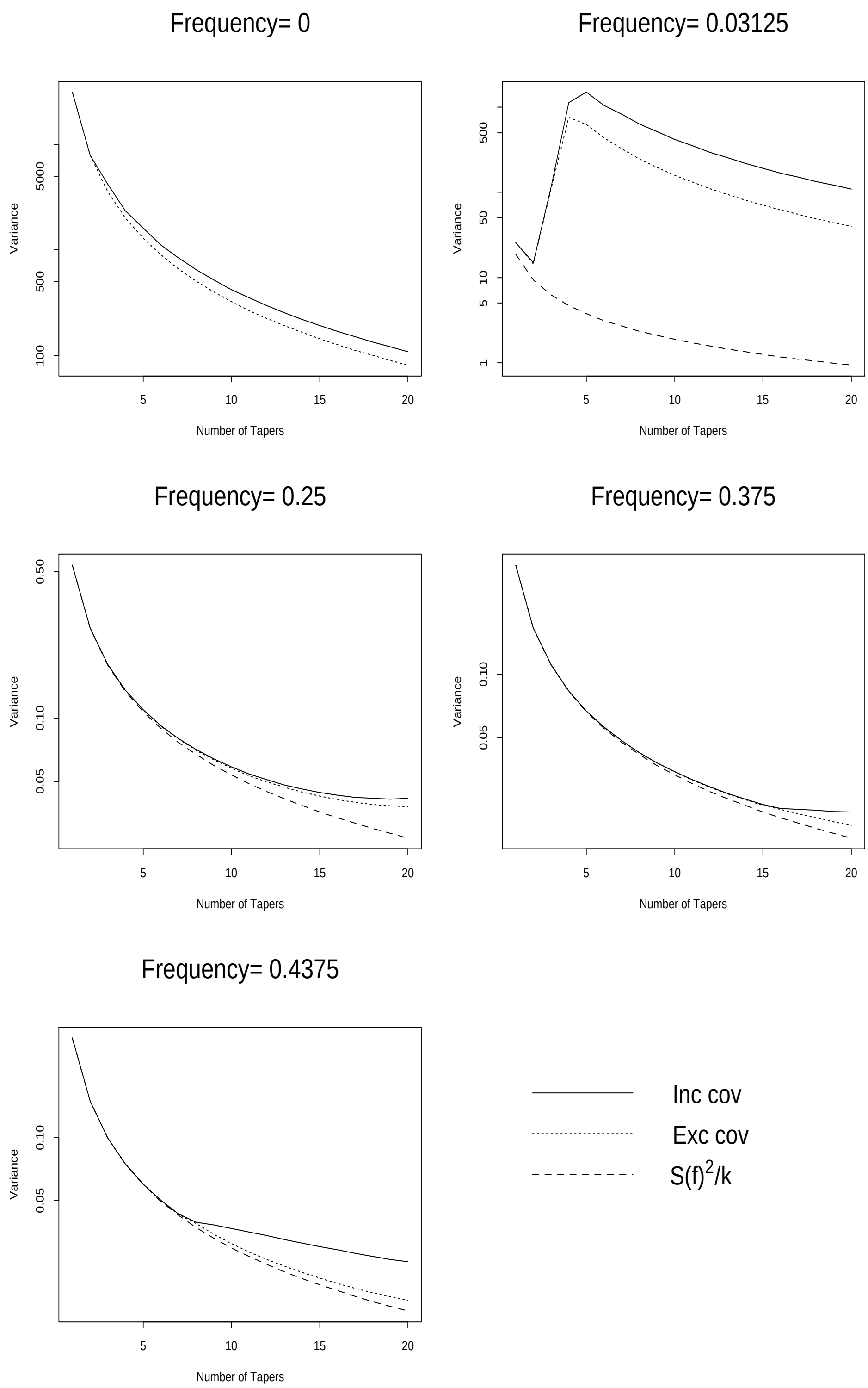


Figure 4.13: Variance of sinusoidal multitaper estimators of an fdp(0.45)

4.4 Investigating the distributional assumptions of multitaper spectral estimators

We can simulate ARFIMA$(0, d, 0)$ processes using Hosking's algorithm, as outlined in Section 2.3, and can then calculate multitaper spectral estimates for this simulated data, thereby gaining some insight into the distributions of the estimates, to ascertain how well they actually follow the theoretical χ^2_{2K} distributions as described in Section 4.3

We are also in a position to evaluate the sample variance of the estimates to establish how closely they agree with our exact method of calculating the variances from the autocovariance sequence (using the algorithm outlined in Table 4.1) .

Assuming independence and letting $X = \hat{S}^{(mt)}(f)$ we know

$$f_X(x) = \frac{2K}{S(f)} \frac{\left(\frac{1}{2}\right)^K}{\Gamma(K)} \left(\frac{2Kx}{S(f)}\right)^{K-1} \exp\left\{-\frac{Kx}{S(f)}\right\},$$

i.e.

$$X \stackrel{d}{=} \frac{S(f)}{2K}\chi^2_{2K} \quad \text{where} \quad S(f) = |2\sin\pi f|^{-2d}, \tag{4.4}$$

where we assume w.l.o.g. that $\sigma^2_\varepsilon = 1$.

Thus, we are in a position to compare our simulation results with the results we would expect under the assumption of independence of the individual direct spectral estimators, which, as we have seen, does not hold for very low frequencies.

4.4.1 Results for fractionally differenced processes

A graphical method of establishing whether a set of data arise from a particular distribution is the Q-Q (Quantile-Quantile) plot, which plots the sorted data against the corresponding quantiles of the chosen distribution. If the points on this plot are approximately linear, then the assumption that the data arise from the distribution is supported, departures from linearity indicate departures in the data from the particular distribution being advocated.

Figures 4.14 to 4.17 show Q-Q plots of the simulated data against the theoretical χ^2 distributions given in Equation 4.4. We see that at low frequencies the theoretical distributions break down for higher numbers of tapers as the value of d

approaches 0.5. As the frequency increases, the theoretical distributions become valid, even for high numbers of tapers.

Reducing the window width W improves the problem of local bias at low frequencies as fewer Fourier frequencies lie within the central lobe, which is the cause of local bias. For comparison, periodogram results are shown in Figures 4.18 to 4.19. The periodogram suffers less from the local bias problem as it has a narrower main lobe (tapering reduces broadband bias at the expense of resolution), however, the periodogram has lower degrees of freedom than a multitaper estimator. Recall that Figures 4.9 and 4.10 show the variances of multitaper estimators for fdp's with $d = 0.25$ and $d = 0.45$ respectively. It is clear from these figures that the variance of the estimators at low frequencies are much higher than would be expected if the estimators were distributed as χ^2_{2K}, indicating that this assumption is invalid when the spectrum varies as rapidly as it does for fdp's at low frequencies when $d = 0.25$ and $d = 0.45$.

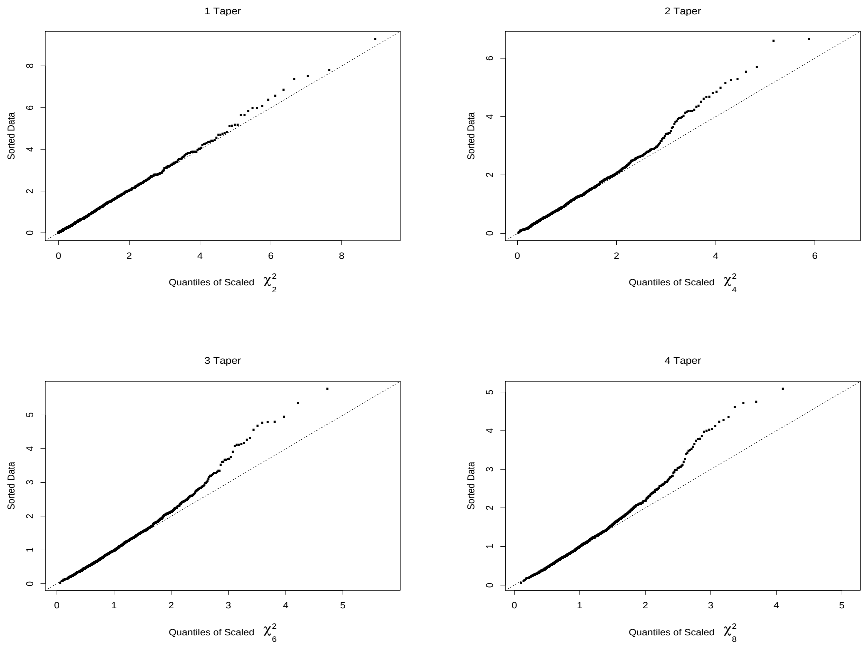


Figure 4.14: Q-Q plots of DPSS ($NW = 4$) multitaper estimators of simulated fdp(0.05) at frequency 0.0625, $N = 64$

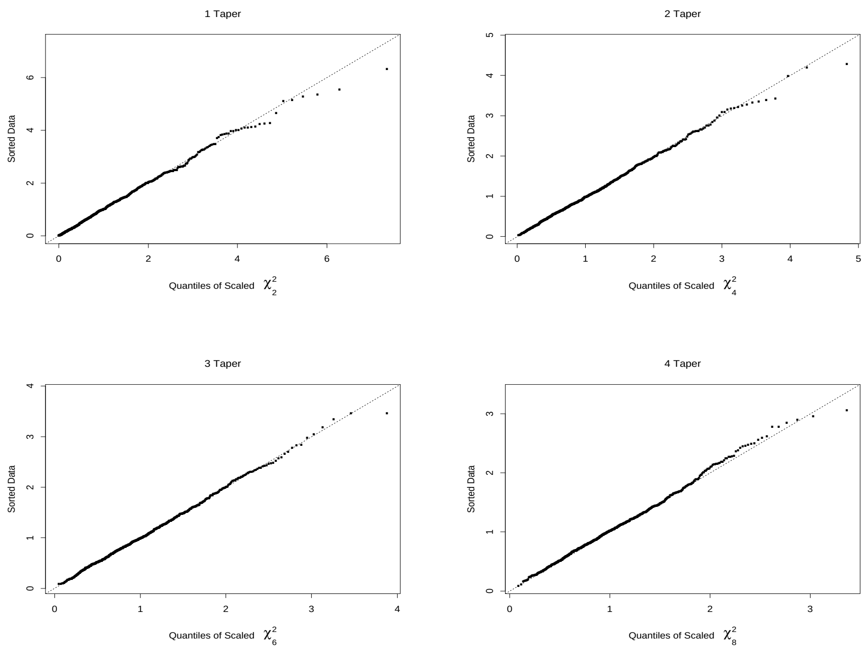


Figure 4.15: Q-Q plots of DPSS ($NW = 4$) multitaper estimators of simulated fdp(0.05) at frequency 0.25, $N = 64$

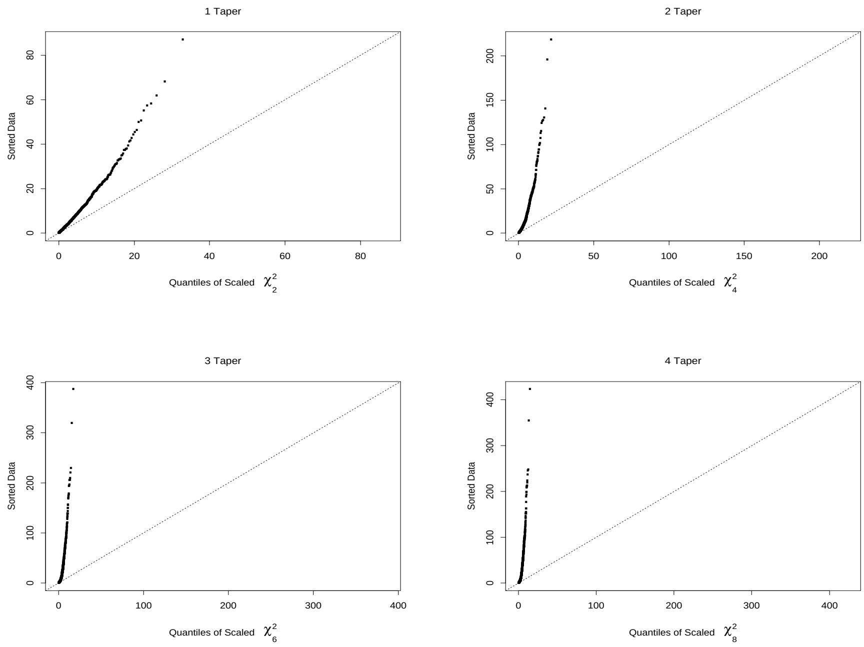


Figure 4.16: Q-Q plots of DPSS ($NW = 4$) multitaper estimators of simulated fdp(0.45) at frequency 0.0625, $N = 64$

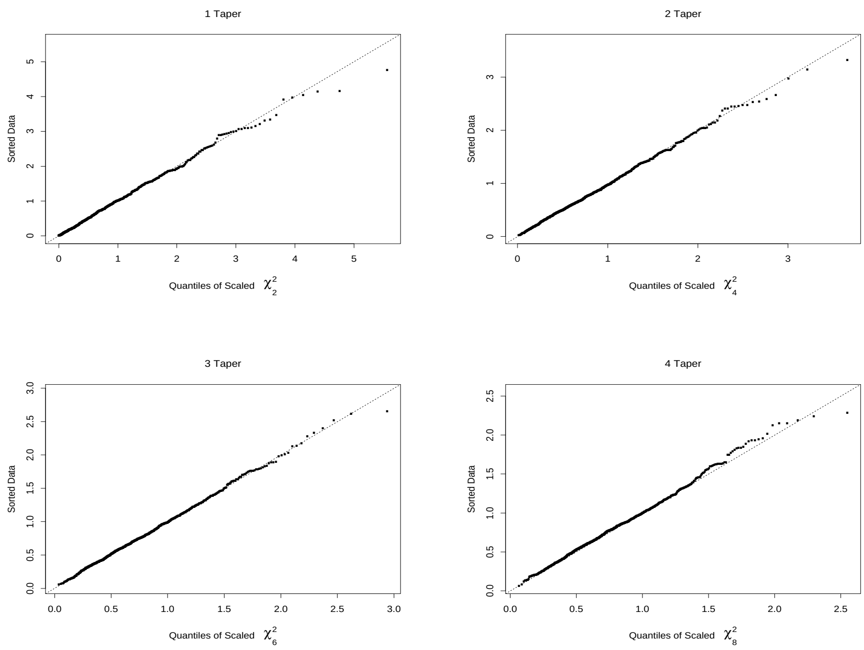


Figure 4.17: Q-Q plots of DPSS ($NW = 4$) multitaper estimators of simulated fdp(0.45) at frequency 0.25, $N = 64$

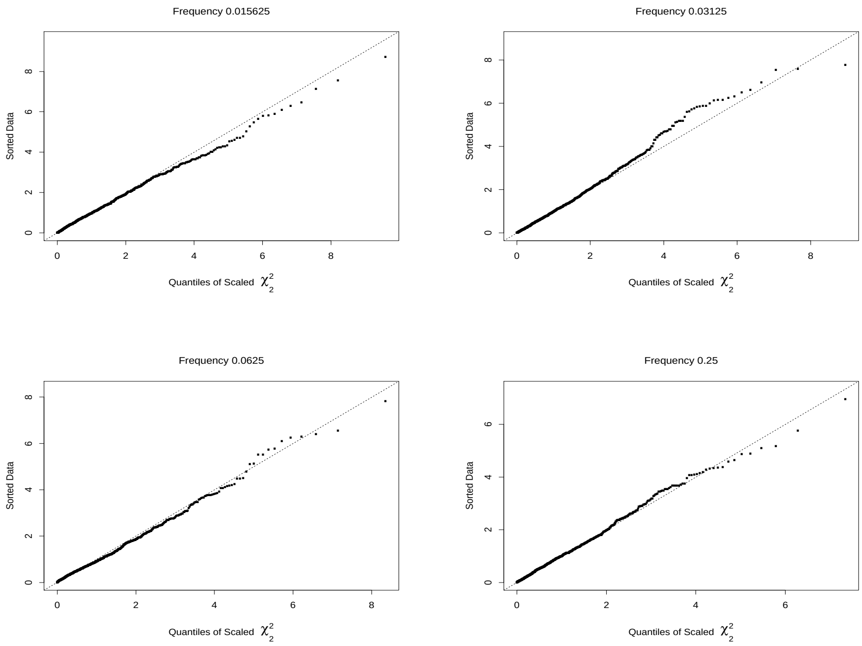


Figure 4.18: Q-Q plots of periodograms of simulated fdp(0.05), $N = 64$

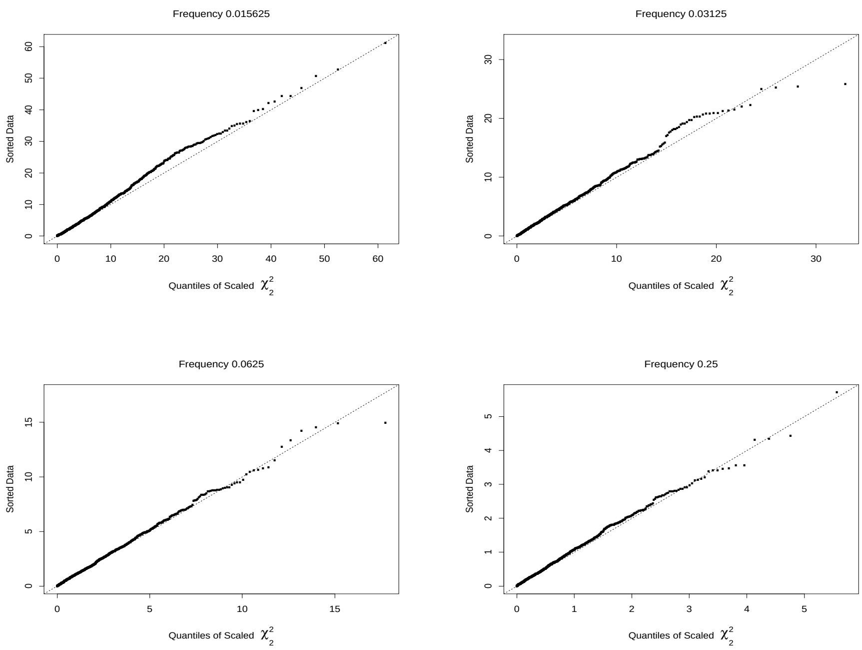


Figure 4.19: Q-Q plots of periodograms of simulated fdp(0.45), $N = 64$

4.5 Effective bandwidth of multitaper spectral estimators

As in the case of all spectral estimators, an important measure is that of the bandwidth of the multitaper estimator. The bandwidth gives us an idea of the frequency domain smoothing, and allows us to establish the distance in frequency between uncorrelated points in the spectral estimator.

In this section we shall extend the bandwidth $\text{width}_a\{\mathcal{H}(\cdot)\}$ described in Equation 3.25 (in Section 3.9) for a direct spectral estimator using a single taper, to the multitaper case (see also Walden, McCoy & Percival (1995)).

The associated spectral window for a multitaper estimator is given by

$$\overline{\mathcal{H}}(f) = \frac{1}{K}\sum_{k=0}^{K-1} \mathcal{H}_k(f),$$

and the equivalent width of its autocorrelation is given by

$$\begin{aligned}\text{width}_a\{\overline{\mathcal{H}}(f)\} &= \frac{\left[\int_{-1/2}^{1/2} \overline{\mathcal{H}}(f)\, df\right]^2}{\int_{-1/2}^{1/2} \overline{\mathcal{H}}^2(f)\, df} \\ &= \frac{\left[\frac{1}{K}\sum_{k=0}^{K-1}\int_{-1/2}^{1/2} \mathcal{H}_k(f)\, df\right]^2}{\int_{-1/2}^{1/2} \overline{\mathcal{H}}^2(f)\, df} \\ &= \left\{\int_{-1/2}^{1/2} \overline{\mathcal{H}}^2(f)\, df\right\}^{-1}.\end{aligned}$$

By Parseval's theorem

$$\int_{-1/2}^{1/2} \overline{\mathcal{H}}^2(f)\, df = \sum_{\tau=-(N-1)}^{(N-1)} \bar{c}_\tau^2,$$

where $\{\bar{c}_\tau\} \longleftrightarrow \overline{\mathcal{H}}(\cdot)$. From which it follows that, assuming that $\Delta t = 1$,

$$\bar{c}_\tau = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{t=1}^{N-|\tau|} h_{t,k}h_{t+|\tau|,k}.$$

Thus

$$\text{width}_a\{\overline{\mathcal{H}}(\cdot)\} = \left\{\sum_{\tau-(N-1)}^{(N-1)}\left[\frac{1}{K}\sum_{k=0}^{K-1}\sum_{t=1}^{N-|\tau|} h_{t,k}h_{t+|\tau|,k}\right]^2\right\}^{-1}. \tag{4.5}$$

Hence, in the case of multitapering, Equation 4.5 can be used to measure the effective bandwidth of the associated spectral window $\overline{\mathcal{H}}(\cdot)$. Referring back to

Figure 4.1, we can see that in the case of the discrete prolate spheroidal sequences, as the number of tapers used in the estimator increases so does the width of the main lobe of $\overline{\mathcal{H}}(\cdot)$, which goes from being bell shaped to approximately rectangular, when the maximum number of tapers are used for that particular design bandwidth. We would expect a similar pattern in any sensible estimator of the bandwidth.

We can see from Figure 4.20, which shows Equation 4.5 calculated for the discrete prolate spheroidal sequences with parameter $NW = 2, 3, 4$ (the shortest line corresponding to $NW = 2$, and the longest line corresponding to $NW = 4$), and $N = 32$, that the effective bandwidth does indeed increase with increasing number of tapers. The dashed lines in the figure show the design bandwidths $2W$. We see that our measure of bandwidth approaches the design bandwidths as the number of tapers used approaches the Shannon number $(2NW - 1)$.

Similarly, for the discrete sinusoidal tapers we can see from Figure 4.2 that the width of the mean spectral window increases as the number of tapers used increases. Figure 4.21 shows Equation 4.5 calculated for the discrete sinusoidal tapers, and again we see that the measure increases as the number of tapers used increases, thus agreeing with the shape of the spectral window. The dashed line shows the nominal bandwidth as given by Riedel & Sidorenko (1995): $(k + 1)/(N + 1)$, where k is the number of tapers used. We see that our measure of bandwidth is very close to this nominal measure.

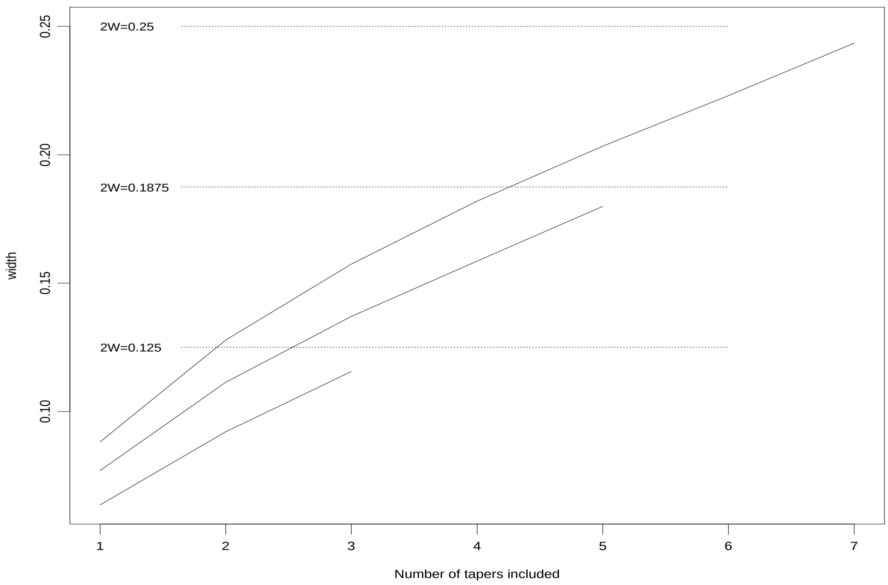


Figure 4.20: $\text{width}_a\{\overline{\mathcal{H}}(\cdot)\}$ for the discrete prolate spheroidal sequences, $NW = 2, 3, 4$ and $N = 32$

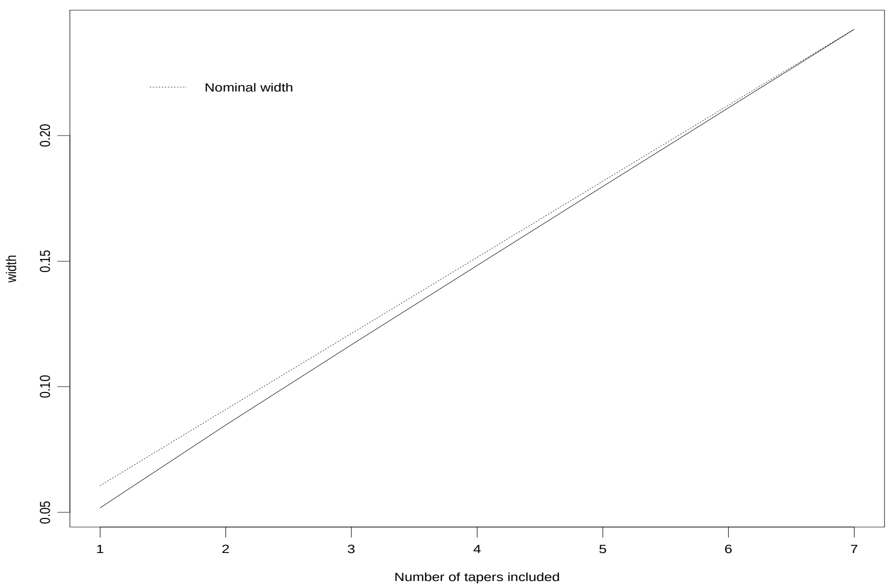


Figure 4.21: $\text{width}_a\{\overline{\mathcal{H}}(\cdot)\}$ for the discrete sinusoidal tapers $N = 32$

Chapter 5

Estimating the Measure of Persistence via Multitapering

5.1 Introduction

If we suspect that a time series of interest exhibits long memory behaviour, or more particularly, that it may be modeled by a fractionally differenced process, we need an accurate method of estimating the measure of persistence, d.

Several authors have considered this problem, generally from two perspectives; firstly by a maximum likelihood method, and secondly by some form of regression on estimates of the log spectrum.

Asymptotic normality of the maximum likelihood estimator has been proved by Yajima (1988) and Dahlhaus (1989). Efficient maximum likelihood estimation has been outlined by several authors including Li & McLeod (1986), Fox & Taqqu (1986), Deriche & Tewfik (1993), and Haslett & Raftery (1989).

Geweke & Porter-Hudak (1983) suggest estimating d via least squares regression, using only the lowest frequency ordinates of the log periodogram, while Kashyap & Eom (1988) suggest a similar method. This is due to the fact that they are estimating the parameter for models which are designed to include both long and short memory, so excluding the high frequency ordinates excludes this short memory behaviour. Smith (1992) suggests a refinement of this method by using maximum likelihood on the ordinates of the log periodogram, rather than least squares, to take into account the non-normality of the log periodogram ordinates. However Agiakloglou *et al* (1993) show that, for extreme values of d (specifically outside the bounds of stationarity), estimators based on the regression method can be badly biased. Fougère (1985) suggests that estimating the spectrum using Burg's maximum entropy method, as opposed to the periodogram, provides improved estimators. Hassler (1993), Reisen (1994) and Chen *et al* (1994) suggest estimating the spectrum using alternative spectral estimation techniques before performing the regression. Beran & Terrin (1994) suggest an estimator based on a multivariate central limit theorem.

Reisen (1994) shows that parameter estimates obtained using the smoothed periodogram employing the Parzen lag window have bias and mean square error approximately half of that obtained using the unsmoothed periodogram. Chen *et al* (1994) compare parameter estimators obtained using the periodogram and lag window spectral density estimators using several different lag windows. They find that generally, the periodogram provides an estimator with overall smallest bias, but the lag window estimators provide estimators with smaller mean square error, demonstrating the trade off between smoothness and resolution that exists between the periodogram and lag window estimators.

Hurvich & Beltrao (1994) suggest an automatic method of selecting how many observations to use in the regression, when the periodogram ordinates are used.

The method that we shall outline here is based on regression of multitaper estimators of the spectrum. We shall see that this method has several appealing attributes; as we increase the number of tapers included in the multitaper method, so the degrees of freedom associated with the theoretical distributions of the estimators increases. The distributions of the log spectral estimators move from a skewed distribution, to one that is more symmetrical and Gaussian (see Figure 4.3, and the discussion in Section 4.3), and so a least squares method based on the log multitaper estimate may be adequate. The computational simplicity of least squares methods is appealing over maximum likelihood methods.

Unlike the Burg maximum entropy method, with multitapering, we have knowledge of the distributional and correlation structure of the spectral estimator. We will compare our method of estimating d to some of the methods outlined above.

If the value of d lies outside the range of stationarity, we can difference the series (see Janacek (1982)) until the value of d lies within $[0, 1/2)$, and then use the methods of approximation outlined here.

New work in this chapter is the refinement of the multitaper method of estimating d first introduced by Percival & Walden (1991). This new method takes into account the problems of smoothing bias encountered by multitaper spectral estimators at low frequencies for fractionally differenced processes.

5.2 Estimating using the estimated spectrum

Percival & Walden (1991) introduce a method of estimating the measure of persistence d, using the logarithm of the multitaper estimate of the spectrum.

Recall, for a fdp we have

$$S(f) = \Delta t \sigma_\varepsilon^2 |2 \sin \pi f|^{-2d} \text{ for } 0 < f < \tfrac{1}{2\Delta t}.$$

Given $\{X_1, \ldots, X_N\}$ and assuming both $\Delta t = 1$ and $\sigma_\varepsilon^2 = 1$, we have

$$\log S(f_j) = -2d \log\{2|\sin \pi f_j|\}, \tag{5.1}$$

where f_j are the Fourier frequencies j/N for $j = 0, \ldots, N/2$. Now, recall that for a multitaper estimate, approximately

$$\frac{\hat{S}^{(mt)}(f_j)}{S(f_j)} \stackrel{d}{=} \frac{1}{2K} \chi^2_{2K}.$$

Adding $\log\{\hat{S}^{(mt)}(f_j)\}$ to both sides of Equation 5.1,

$$\log\{\hat{S}^{(mt)}(f_j)\} = -2d \log(2|\sin \pi f|) + \log\left\{\frac{\hat{S}^{(mt)}(f_j)}{S(f_j)}\right\}.$$

So, we are in a position to estimate d using linear regression

$$y_j = \alpha x_j + \varepsilon_j, \quad j = 1, \ldots, N/2,$$

where:

$$\begin{aligned}
y_j &= \log\{\hat{S}^{(mt)}(f_j)\} \\
x_j &= \log\{2\sin|\pi f_j|\} \\
\varepsilon_j &= \log\{\hat{S}^{(mt)}(f_j)/S(f_j)\} \\
\alpha &= -2d.
\end{aligned}$$

Here the dependent variable is $y_j = \log\{\hat{S}^{(mt)}(f_j)\}$ and the independent variable is $x_j = \log\{2\sin|\pi f_j|\}$. The distribution of the $\{\varepsilon_j\}$ is given by

$$\varepsilon_j \stackrel{d}{=} \log\left\{\frac{1}{2K}\chi^2_{2K}\right\}.$$

Then

$$E\{\varepsilon_j\} = \psi(K) - \log(K) \quad \text{and} \quad \text{var}\{\varepsilon_j\} = \psi'(K),$$

where ψ and ψ' are, respectively, the gamma and digamma functions. We investigated the distribution of $\{\varepsilon_j\}$ in Section 4.3, and saw that, as the degrees of freedom increases, so the distribution becomes more Guassian, indicating that least square estimation will be close to optimal.

If we rewrite the regression as

$$y_j = c + \alpha x_j + \eta_j, \quad j = 1, \ldots, N/2,$$

where $c = \psi(K) - \log(K)$ and $\eta_j = \varepsilon_j - \psi(K) + \log(K)$, then $E\{\eta_j\} = 0$ and $\text{var}\{\eta_j\} = \psi'(K)$. Then estimating d by least-squares regression leads to the estimator

$$\hat{d} = -\frac{\sum_{i=1}^{N/2}(x_i - \bar{x})y_i}{2\sum_{i=1}^{N/2}(x_i - \bar{x})^2},$$

using the distribution of the $\{\varepsilon_j\}$ leads to $E(\hat{d}) = d$ and

$$\text{var}(\hat{d}) = \frac{\psi'(K)}{4\sum_{i=1}^{N/2}(x_i - \bar{x})^2}. \tag{5.2}$$

The variance of the estimator $\hat{d}_p$ obtained from regression based on the periodogram, is given by:

$$\text{var}(\hat{d}_p) = \frac{\pi^2}{24\sum_{i=1}^{N/2}(x_i - \bar{x})^2}, \tag{5.3}$$

e.g. Reisen (1994).

These theoretical variances assume that the ε_j are uncorrelated. In Section 3.8, we investigated the correlations that exist between direct spectral estimators across frequencies. We showed that the periodogram is uncorrelated at the Fourier frequencies, but that spectral estimators whose associated tapers have a spectral window with a wider main lobe that Fejér's kernel, exhibit correlations at the Fourier frequencies. These correlations are inherited by the multitaper estimators. The implication of these correlations is that the multitaper regression model we have outline above, is not a standard regression model, as the ε_j exhibit correlations, though these correlations may be small.

Table 5.1 compares the theoretical standard deviations, derived from Equations 5.2 and 5.3, of the estimators obtained when using the multitaper and periodogram method. We see that, as the number of tapers used increases, so the variance of the estimator decreases, and is consistently smaller than the periodogram variance. This is due to the fact that $\psi'(K)/4 < \pi^2/24$ when $K > 1$.

	N	
	128	256
Periodogram	0.10056	0.06769
1 taper	0.10056	0.06769
2 taper	0.06296	0.04239
3 taper	0.04927	0.03317
4 taper	0.04177	0.02812
5 taper	0.03689	0.02483
6 taper	0.03339	0.02247
7 taper	0.03072	0.02068

Table 5.1: Comparison of theoretical standard deviations of estimators

Although we have outlined the method using all the Fourier frequencies in the regression, in practice we do not use all the points in the regression, but exclude those that are closer to zero frequency than the estimated bandwidth of the mean spectral window which we developed in Section 4.5. We exclude these points to ensure that the results are not biased by the large concentration of energy at very low frequencies. This is in contrast to the methods of Reisen (1994) and Chen *et al* (1994), who exclude high frequency points in the regression as they are investigating models including short memory behaviour, which they wish to exclude from their analyses.

We shall compare this multitaper method of estimating d to the other methods previously suggested, that were outlined in the introduction.

5.3 Simulation results: comparison with periodogram

We have simulated 500 fractionally differenced datasets using Hosking's algorithm which was described in Table 2.1, Section 2.3, for dataset lengths of 128 and 256 and values of d of 0.25 and 0.4, and $\sigma_\epsilon^2 = 1$. We have then estimated d from regression of the multitaper estimates as outlined in the previous section, and for comparison, we have also estimated d using regression of the periodogram including all frequency ordinates in the regression. Table 5.2 shows the results obtained for the 5 taper estimates using both the discrete prolate spheroidal sequences with $NW = 4$, and the discrete sinusoidal tapers. We show the mean estimate of d over the 500 simulations, and we also show estimates of the bias and mean square error which we estimate as follows:

$$\text{bias} = \frac{1}{500}\sum_{j=1}^{500}(\hat{d}_j - d)$$

and

$$\text{mean square error} = \frac{1}{500}\sum_{j=1}^{500}(\hat{d}_j - d)^2.$$

The tables also include the estimated standard deviations, so that we can compare them to the theoretical results shown in Table 5.1.

We can see from this table that the periodogram estimates of d tend to be less biased than the multitaper estimates, but have a considerably larger mean square error. The sinusoidal tapers perform better than the discrete prolate spheroidal sequences in these simulations. An explanation for this could be that the bandwidth of the mean spectral window when using the discrete sinusoidal tapers is less than that when using discrete prolate spheroidal sequences (at least up to the Shannon number).

Our periodogram results produce estimators with lower bias and mean square error than those published in either Reisen (1994) or Chen *et al* (1994), this is due to the fact that these authors do not include all frequency ordinates in the regression, but exclude the higher frequency ordinates. However, our periodogram results do agree with the theoretical standard deviations of the estimators given in Table 5.1, derived from Equation 5.3. The standard deviations for the multitaper estimates are higher than is suggested by Table 5.1 (derived from Equation 5.2), this could be due to the correlations that exist between the points used in the regression.

To compare our results we use boxplots of the simulations. The boxplots shown here are defined as follows: a horizontal line is drawn through the box at the median of the data, the upper and lower ends of the box are at the upper and lower quartiles (so the box covers half the data), and the vertical lines extend from the box to points within a standard range of the data, where the standard range is defined as $1.5\star$(inter-quartile range). Points that are very extreme are plotted by themselves.

Figures 5.1 to 5.8 show boxplots of the results obtained from these simulations, which also show the results obtained using regression of the periodogram. The horizontal line shows the true value of d. We can see that the single taper estimate provides estimates of d with larger variation than the periodogram, but as the number of tapers included in the estimate increases, so the estimate of d improves. However the bias in the estimate also appears to positively increase as the number of tapers increases, this could be due to the increase in bandwidth of the mean spectral window.

d		Periodogram		DPSS's		Sine tapers	
		128	256	128	256	128	256
0.25	mean	0.2636	0.2548	0.2717	0.2555	0.2738	0.2549
	bias	0.0136	0.0048	0.0217	0.0055	0.0238	0.0049
	mse	0.0108	0.0046	0.0076	0.0027	0.0068	0.0025
	std dev	0.1030	0.0681	0.0843	0.0515	0.0790	0.0496
0.4	mean	0.3879	0.3969	0.4181	0.4092	0.4113	0.4063
	bias	0.0176	-0.0031	0.0181	0.0092	0.0113	0.0063
	mse	0.0107	0.0045	0.0075	0.0018	0.0048	0.0019
	std dev	0.1179	0.0668	0.0848	0.0418	0.0686	0.0436

Table 5.2: Comparison of estimates of d obtained from multitaper spectral estimates and the periodogram

5.4 Simulation results: comparison with Burg's maximum entropy spectral estimation

In this section we compare the multitaper results with those obtained using the method outline in Fougère (1985), which employs Burg's maximum entropy method (MEM) of spectral estimation (Burg (1967)).

5.4.1 The method

A stationary AR(p) (autoregressive process of order p) process $\{X_t\}$, with zero mean, satisfies the equation

$$X_t = \phi_{1,p}X_{t-1} + \phi_{2,p}X_{t-2} + \ldots + \phi_{p,p}X_{t-p} + \epsilon_t, \tag{5.4}$$

where $\phi_{j,p}$ for $j = 1, \ldots, p$ are p fixed coefficients, and the $\{\epsilon_t\}$ is the *innovations process* with *innovations variance* σ_p^2. We now look at a well known method of estimating the parameters of this model.

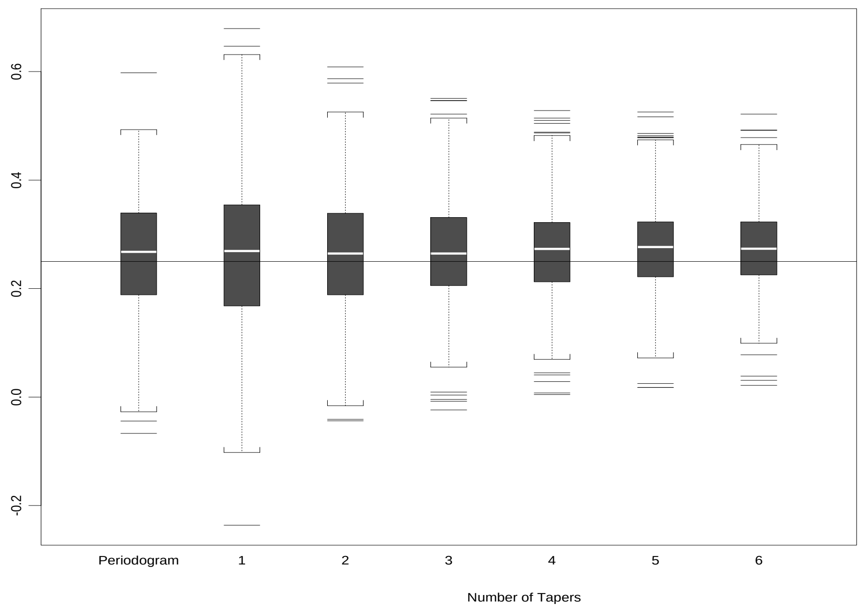


Figure 5.1: Estimates of d, $N = 128$, using DPSS's, $d = 0.25$

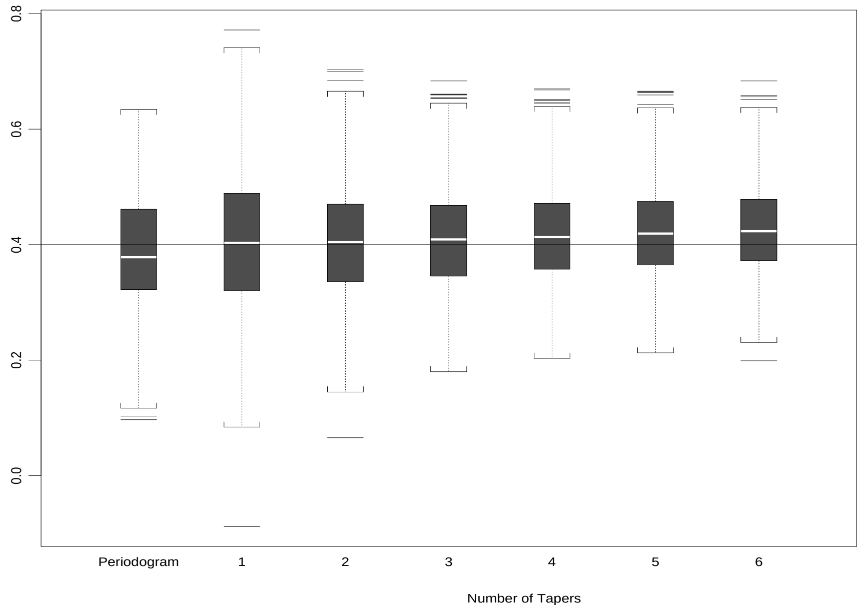


Figure 5.2: Estimates of d, $N = 128$, using DPSS's, $d = 0.4$

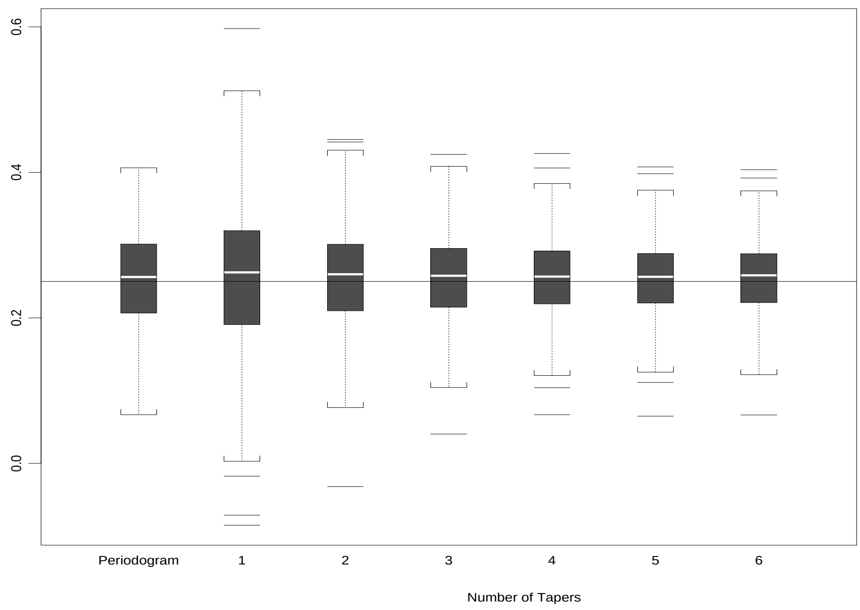


Figure 5.3: Estimates of d, $N = 256$, using DPSS's, $d = 0.25$

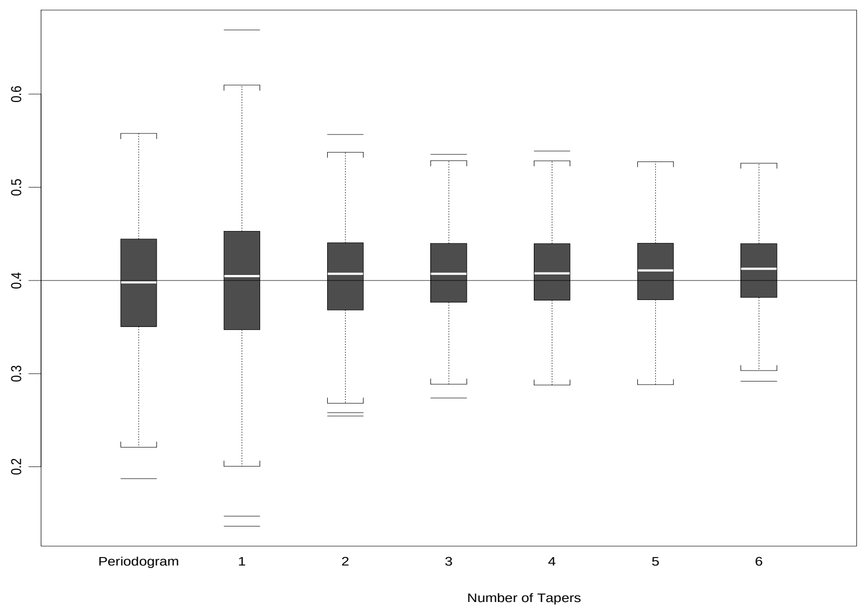


Figure 5.4: Estimates of d, $N = 256$, using DPSS's, $d = 0.4$

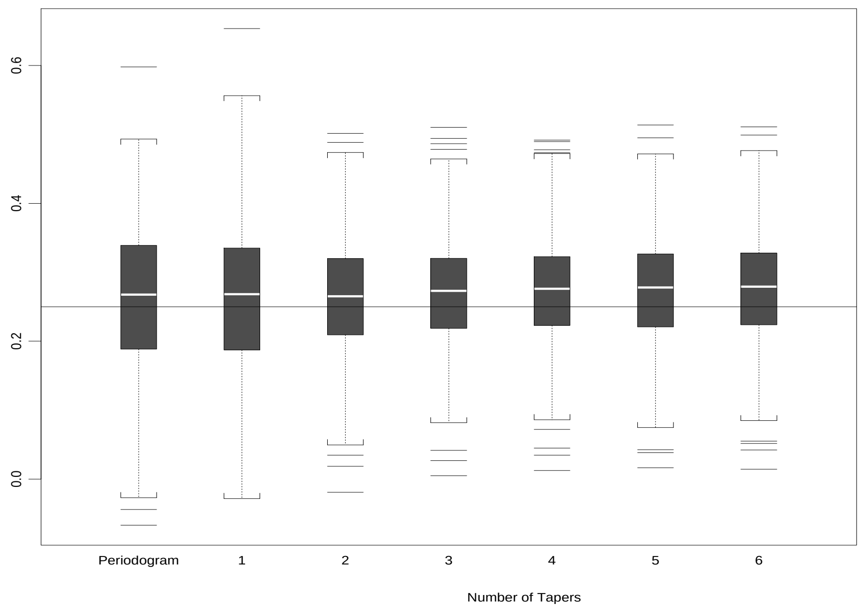


Figure 5.5: Estimates of d, $N = 128$, using discrete sinusoidal tapers, $d = 0.25$

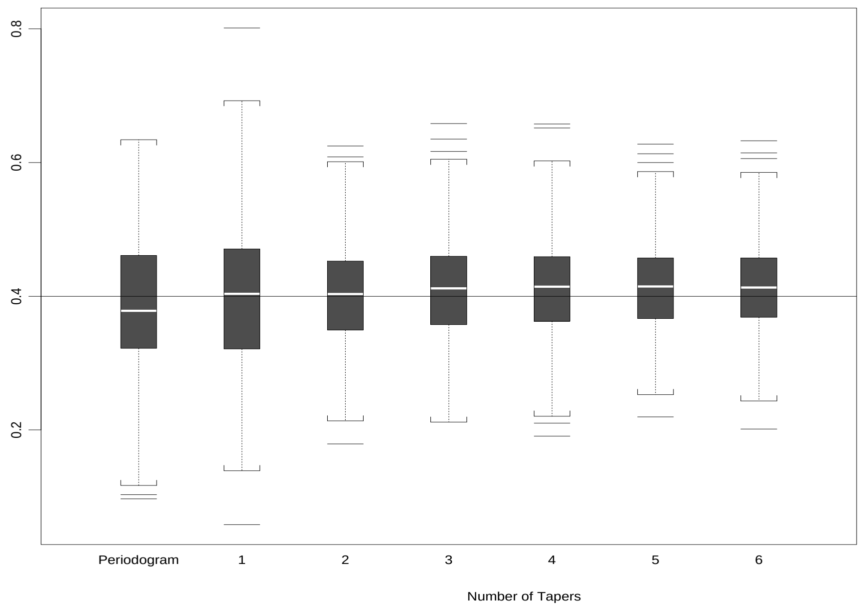


Figure 5.6: Estimates of d, $N = 128$, using discrete sinusoidal tapers, $d = 0.4$

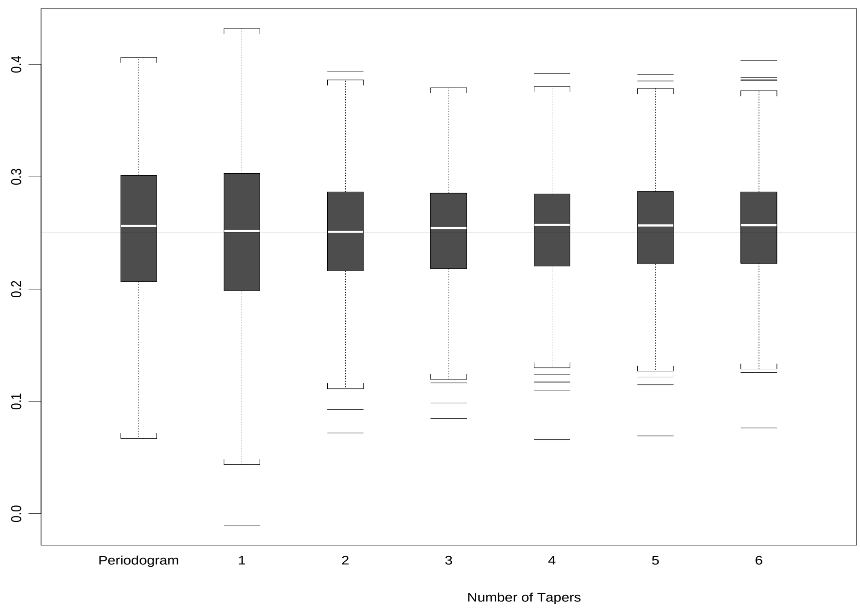


Figure 5.7: Estimates of d, $N = 256$, using discrete sinusoidal tapers, $d = 0.25$

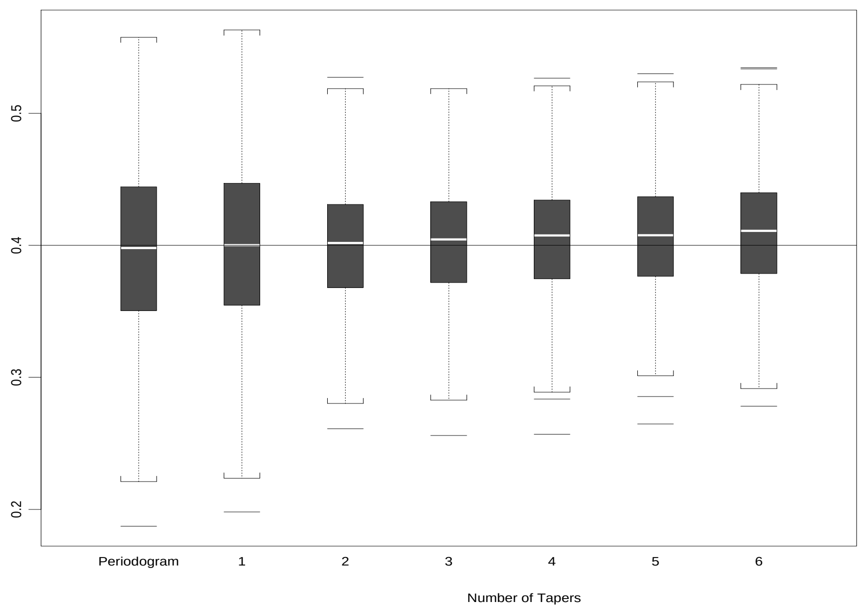


Figure 5.8: Estimates of d, $N = 256$, using discrete sinusoidal tapers, $d = 0.4$

The Yule-Walker equations

The oldest method of estimating the parameters of an AR(p) process $\{X_t\}$ with zero mean and acvs given by $s_k = E\{X_t X_{t+k}\}$ is by matching lagged moments. Multiplying both sides of Equation 5.4 by X_{t-k}:

$$X_t X_{t-k} = \sum_{j=1}^{p} \phi_{j,p} X_{t-j} X_{t-k} + \epsilon_t X_{t-k}.$$

Taking expectations of both sides, we have, for $k > 0$,

$$s_k = \sum_{j=1}^{p} \phi_{j,p} s_{k-j},$$

as $E\{\epsilon_t X_{t-k}\} = 0$. Let $k = 1, 2, \ldots, p$ in the previous Equation and using the fact that $s_{-j} = s_j$ results in the *Yule-Walker equations*:

$$\begin{array}{ccccccccc}
s_1 & = & \phi_{1,p} s_0 & + & \phi_{2,p} s_1 & + & \ldots & + & \phi_{p,p} s_{p-1} \\
s_2 & = & \phi_{1,p} s_1 & + & \phi_{2,p} s_0 & + & \ldots & + & \phi_{p,p} s_{p-2} \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
s_p & = & \phi_{1,p} s_{p-1} & + & \phi_{2,p} s_{p-2} & + & \ldots & + & \phi_{p,p} s_0.
\end{array}$$

We will now outline a recursive method for solving the Yule-Walker equations.

Levinson-Durbin Recursion

There is a recursive method for obtaining estimates for a kth order model from the estimates of the $k-1$ model known as the Levinson-Durbin recursion (e.g. Box & Jenkins (1976, p84)), which recursively solve the *Yule-Walker equations*. If we have estimates $\hat{\phi}_{1,k} \ldots \hat{\phi}_{k-1,k-1}$ and $\hat{\sigma}^2_{k-1}$, assuming an AR($k-1$) model, then estimates for an AR(k) model are

$$\hat{\phi}_{k,k} = \frac{\hat{s}^{(p)}_k - \sum_{j=1}^{k-1} \hat{\phi}_{j,k-1} \hat{s}^{(p)}_{k-j}}{\hat{\sigma}^2_{k-1}} \tag{5.5}$$

$$\hat{\phi}_{j,k} = \hat{\phi}_{j,k-1} - \hat{\phi}_{k,k} \hat{\phi}_{k-j,k-1}, \quad 1 \leq j \leq k-1; \tag{5.6}$$

$$\hat{\sigma}^2_k = \hat{\sigma}^2_{k-1}(1 - \hat{\phi}^2_{k,k}), \tag{5.7}$$

where $\hat{s}^{(p)}_k$ is the estimated autocovariance sequence:

$$\hat{s}^{(p)}_k = \frac{1}{N} \sum_{t=1}^{N-|k|} X_t X_{t+|k|},$$

and the recursion is initialized by

$$\hat{\phi}_{1,1} = \hat{s}_1^{(p)} / \hat{s}_0^{(p)} \quad \text{and} \quad \hat{\sigma}_1^2 = \hat{s}_0^{(p)} - \hat{\phi}_{1,1}\hat{s}_1^{(p)} = \hat{s}_0^{(p)}(1 - \hat{\phi}_{1,1}^2).$$

Burg's algorithm

Burg's algorithm is an alternative method of solving the Yule-Walker equations via the Levinson-Durbin recursions. In the Levinson-Durbin recursions: if we have estimates $\bar{\phi}_{1,k-1}, \ldots \bar{\phi}_{k-1,k-1}$ and $\bar{\sigma}_{k-1}^2$, we can see, using equations analogous to Equations 5.6 and 5.7, that we only need $\bar{\phi}_{k,k}$ in order to calculate estimates of the remaining order k terms (that is $\bar{\phi}_{1,k}, \ldots \bar{\phi}_{k-1,k}$ and $\bar{\sigma}_k^2$). In the Yule-Walker scheme, Equation 5.5 is used to calculate the estimate of $\phi_{k,k}$. Burg's algorithm employs a different method of estimating $\phi_{k,k}$ by minimising a certain sum of squares of *observed* forward and backward prediction errors:

$$\mathrm{SS}_k(\bar{\phi}_{k,k}) \equiv \sum_{t=k+1}^{N} \left[\left\{ X_t - \sum_{j=1}^{k} \bar{\phi}_{j,k} X_{t-j} \right\}^2 + \left\{ X_{t-k} - \sum_{j=1}^{k} \bar{\phi}_{j,k} X_{t-k+j} \right\}^2 \right].$$

This gives rise to a new estimator of $\phi_{k,k}$ which can then be used in the recursion. The recursion is initialised by finding the value of $\bar{\phi}_{1,1}$ that minimises $\mathrm{SS}_1(\bar{\phi}_{1,1})$, and by setting $\bar{\sigma}_0^2 = \hat{s}_0^{(p)}$ (the sample variance).

Burg's maximum entropy spectral estimator

The parameter estimators obtained using Burg's algorithm described in the previous section, are then used in the equation for the theoretical spectrum of an AR(p) process, to obtain the Burg maximum entropy spectral estimator:

$$\hat{S}(f) = \frac{\bar{\sigma}_p^2}{\left|1 - \sum_{j=1}^{p} \bar{\phi}_{j,p} e^{-i2\pi f j}\right|^2}.$$

This algorithm is optimal in the sense of maximising a measure of entropy, though can be justified as above without appealing to entropy arguments. Burg's maximum entropy spectral estimation has many good properties: it provides a continuous function of frequency, and has been shown to produce a smooth spectrum while maintaining high resolution. However, one must choose the order of the AR model with care, and for long memory processes, automatic methods such as the Akaike information criterion, may underestimate the number of parameters required to capture the true spectrum close to zero. Burg's maximum entropy method also has no associated distribution in the way that the multitaper estimator does, so we can draw no theoretical conclusion about the regression parameters. For these simulations, the number of parameters in the model have been fixed at 6, following Fougère (1985). We have used the S-PLUS time series function `ar.burg` to estimate the AR parameters, and the innovations variance. We have used the same datasets as in the previous section, so that we can draw comparisons between all the methods employed.

5.4.2 Results

Table 5.3 shows the results as compared to the 5 taper multitaper results, we see that for the lower values of d and N the methods produce comparable results, but that the Burg maximum entropy method has a larger bias for the higher value of d, and this bias does not improve for larger sample sizes.

These results are in contrast to Fougère (1985), who finds from his simulations that the parameter estimators obtained using the Burg maximum entropy method are less biased than the periodogram estimators. The data that Fougère used in his simulations was generated by passing white noise through an infinite impulse response filter which produces a power law spectrum over a limited frequency range. Fougère points out that such a filter would represent an autoregressive process. Thus, the very nature of Burg's MEM is best suited to analyzing the data he generates, and is able to produce very accurate spectral estimates. Here, we have generated our data using the time-domain method devised by Hosking, and so give none of the methods an advantage over any other method.

Figures 5.9 and 5.10 shows boxplots of these results as compared to both the periodogram and the 5 taper multitaper estimates, here we can clearly see the differences between the estimates.

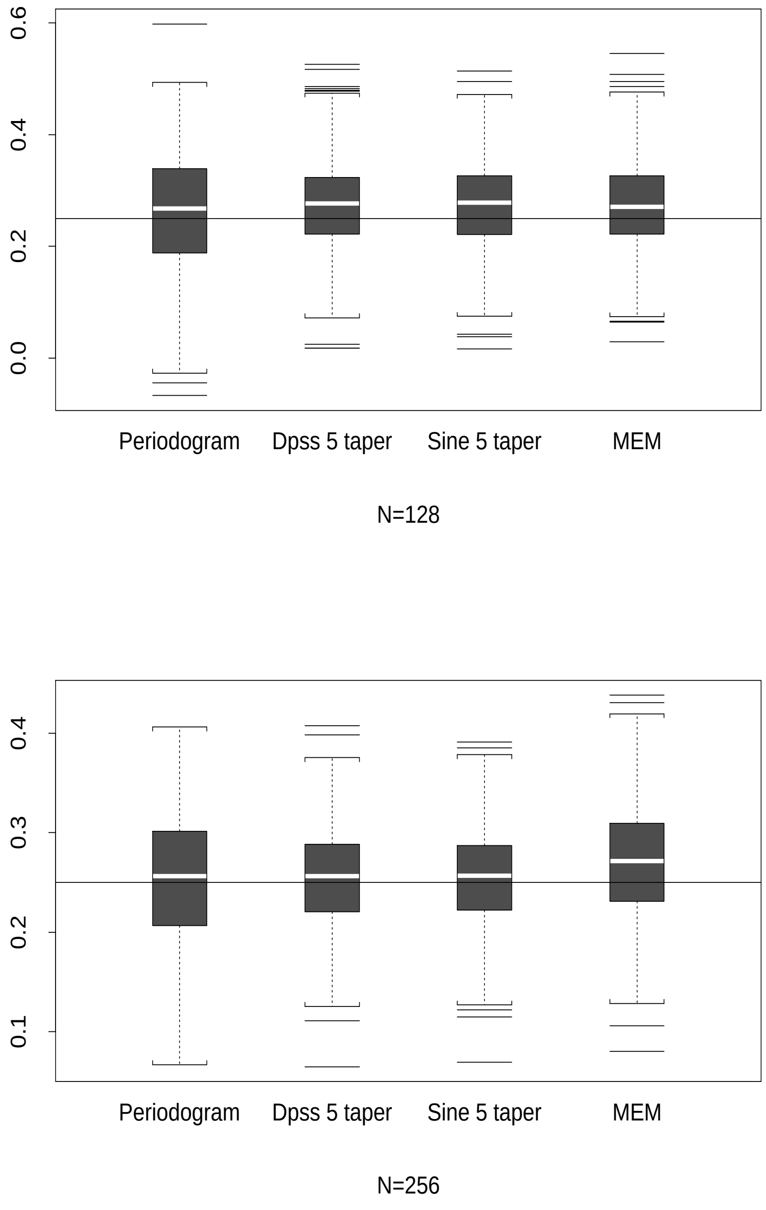
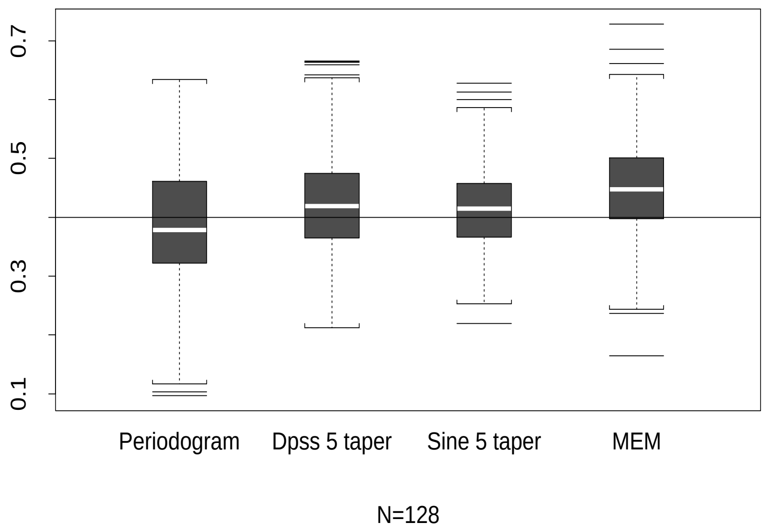


Figure 5.9: Comparison with Burg's maximum entropy method, $d = 0.25$



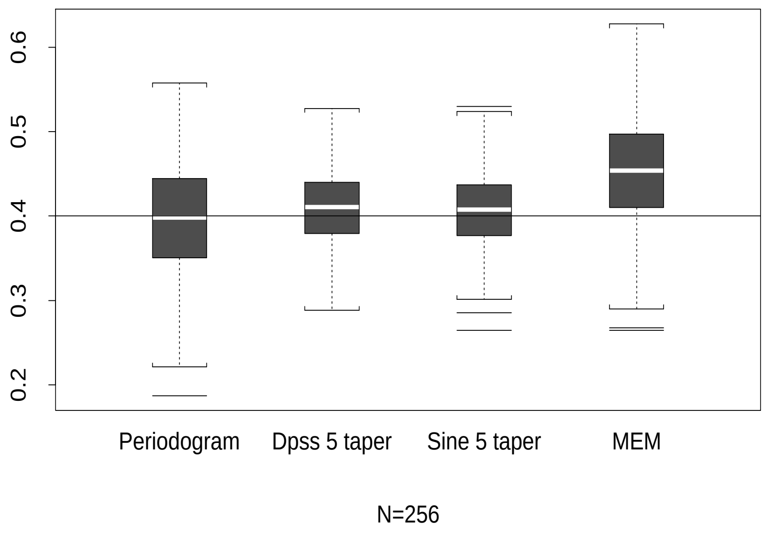


Figure 5.10: Comparison with Burg's maximum entropy method, $d = 0.4$

d		MEM		DPSS's		Sine tapers	
		128	256	128	256	128	256
0.25	mean	0.2730	0.2701	0.2717	0.2555	0.2738	0.2549
	bias	0.0230	0.0201	0.0217	0.0055	0.0238	0.0049
	mse	0.0070	0.0037	0.0076	0.0027	0.0068	0.0025
0.4	mean	0.4462	0.4510	0.4181	0.4092	0.4113	0.4063
	bias	0.0462	0.0510	0.0181	0.0092	0.0113	0.0063
	mse	0.0087	0.0065	0.0075	0.0018	0.0048	0.0019

Table 5.3: Comparison of estimates of d obtained from multitaper spectral estimates and Burg's maximum entropy method

5.5 Simulation results: comparison with maximum likelihood

In this section we compare multitaper results with those obtained using the fast approximate maximum likelihood algorithm of Haslett & Raftery (1989). This algorithm carries out a univariate unimodal optimization with respect to d over the interval $[0, 0.5)$, we shall now describe this method in more detail.

5.5.1 Maximum likelihood algorithm

Haslett & Raftery's algorithm is designed to estimate the parameters of a full ARFIMA(p, d, q) model. Here we will describe their algorithm when applied to the more specialised fractionally differenced white noise models – ARFIMA$(0, d, 0)$, that concern us here.

Recall from Section 2.3, that a Gaussian stationary linear stochastic process $\{X_t\}$ with zero mean, will have multivariate normal distribution with conditional mean and variance of X_t, given the past values $X_{t-1}, X_{t-2}, \ldots, X_0$, given by

$$\begin{aligned} m_t &\equiv E(X_t|X_{t-1}, X_{t-2}, \ldots, X_0) = \sum_{j=1}^{t} \phi_{tj} X_{t-j}, \\ v_t &\equiv \operatorname{var}(X_t|X_{t-1}, X_{t-2}, \ldots, X_0) = s_0 \prod_{j=1}^{t} (1 - \phi_{jj}^2), \end{aligned}$$

(Ramsey (1974)), where $s_0 = \text{var}(X_t)$, ϕ_{ii} is the ith partial correlation coefficient of $\{X_t\}$, and the $\{\phi_{ti}\}$ are the partial linear regression coefficients, which can be calculated using the Levinson-Durbin recursion (Box & Jenkins (1976, p.84)). The Haslett & Raftery algorithm calculates these partial linear regression coefficients exactly up to some integer M, and approximates them by $\phi_{ti} \approx -\pi_i$ for $i > M$ (Hosking (1981)), where π_i are the π weights of the ARFIMA$(0, d, 0)$ process, as defined in Box & Jenkins (1976). Given a value of d, the approximate likelihood estimator of σ_ϵ^2 can then be derived analytically.

Using an initial estimator of d, we can calculate an estimator of σ_ϵ^2, and form the concentrated log-likelihood by replacing σ_ϵ^2 with its estimate. An approximate maximum likelihood estimator of d can then be found, by maximising this concentrated log-likelihood numerically.

5.5.2 Results

The results for the data used in the previous sections are given in Table 5.4, and here we can see that while the mean square error of the estimates obtained with the maximum likelihood method are comparable with the multitaper estimates, the estimates suffer from a significant negative bias.

d		Maximum Likelihood		DPSS's		Sine tapers	
		128	256	128	256	128	256
0.25	mean	0.2049	0.2285	0.2717	0.2555	0.2738	0.2549
	bias	-0.0451	-0.0215	0.0217	0.0055	0.0238	0.0049
	mse	0.0077	0.0032	0.0076	0.0027	0.0068	0.0025
0.4	mean	0.3433	0.3725	0.4181	0.4092	0.4113	0.4063
	bias	-0.0567	-0.0275	0.0181	0.0092	0.0113	0.0063
	mse	0.0082	0.0031	0.0075	0.0018	0.0048	0.0019

Table 5.4: Comparison of estimates of d obtained from multitaper spectral estimates and approximate maximum likelihood

Figures 5.11 and 5.12 shows boxplots of these results as compared to both the periodogram and the 5 taper multitaper estimates, here we can clearly see the bias in the approximate maximum likelihood estimates.

5.6 Analysis of geophysical borehole data

In this section we investigate the power law spectra exhibited by the reflection coefficients from a variety of rock sequences. The data we show here are derived from sonic logs from three of eight wells studied by Walden & Hosken (1985). Walden & Hosken conclude the following for these three wells:

Well 1: slope of spectrum of reflection coefficients $\propto f$

Well 2: slope of spectrum of reflection coefficients $\propto f^{1/2}$

Well 3: slope of spectrum of reflection coefficients $\propto f^{3/2}$

The reflection coefficients for the three wells are shown in Figure 5.13.

The reflection coefficients - $\{r_t\}$, measure the difference in the acoustic impedance (velocity times density) between layers of rock. When the reflection coefficients are small, we can derive an approximation to the actual impedances - $\{I_j\}$ of each layer, using the following

$$\log \frac{I_{j+1}}{I_1} \approx \sum_{t=1}^{j} r_t \quad k = 1, 2, \ldots.$$

The following relationship holds between the reflection coefficients and the logarithmic acoustic impedance:

$$\begin{tikzcd}[row sep=huge] \textbf{Reflection Coefficients} \arrow[d, shift right=6, "\text{Integrate } (\div f^2)"'] \\ \textbf{Logarithmic Acoustic Impedance} \arrow[u, shift right=6, "\text{Differentiate } (\times f^2)"'] \end{tikzcd}$$

From the results of Walden & Hosken, we should have the following results for the logarithmic acoustic impedance:

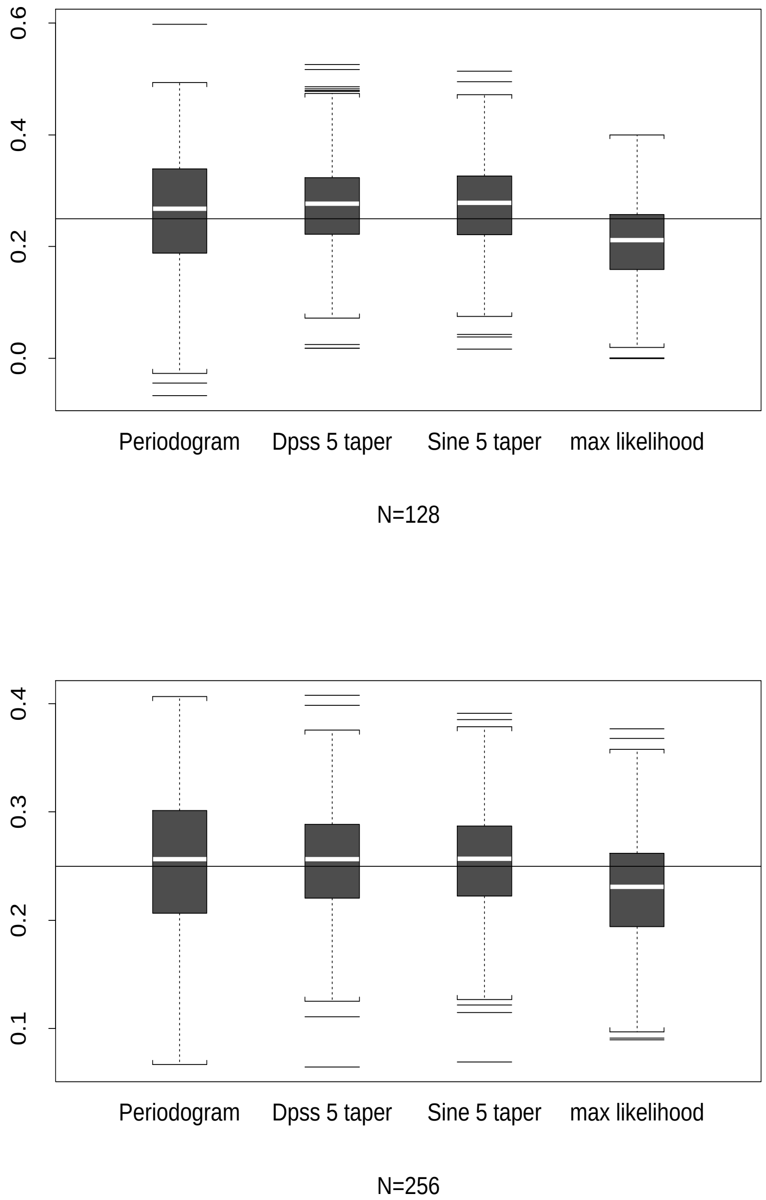


Figure 5.11: Comparison with Haslett & Raftery's approximate maximum likelihood estimates, $d = 0.25$

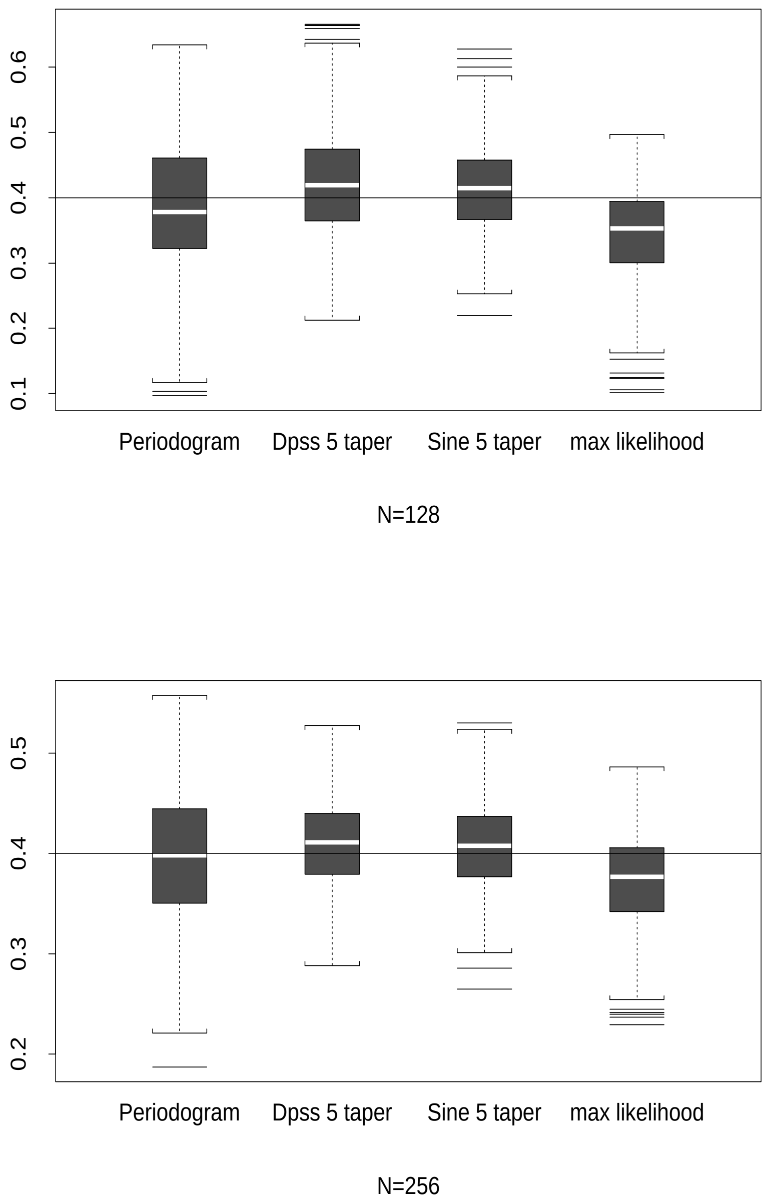


Figure 5.12: Comparison with Haslett & Raftery's approximate maximum likelihood estimates, $d = 0.4$

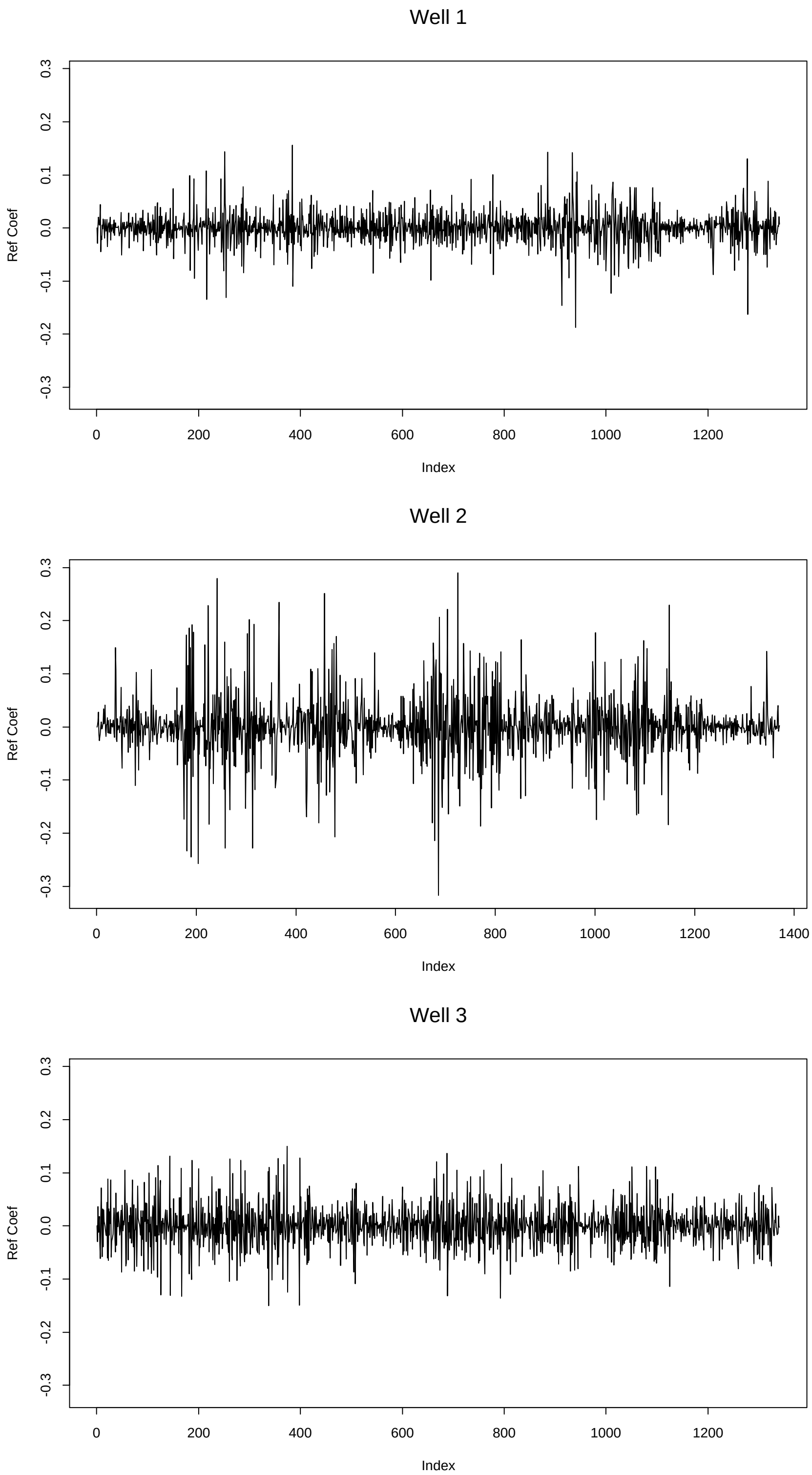


Figure 5.13: Reflection coefficients

Well 1: slope of spectrum of log. acoustic impedance $\propto \frac{1}{f}$.

Well 2: slope of spectrum of log. acoustic impedance $\propto \frac{1}{f^{3/2}}$.

Well 3: slope of spectrum of log. acoustic impedance $\propto \frac{1}{f^{1/2}}$.

The values of logarithmic acoustic impedance will decrease as depth increases, resulting in a positive trend in the derived series. We have removed this trend by assuming that it is linear, and simply subtracting the fitted straight line from the data. The detrended derived values of the logarithmic acoustic impedance for the three wells are shown in Figure 5.14.

For a fractionally differenced process

$$S(f) \approx \frac{1}{f^{2d}} \text{ as } f \to 0.$$

So, we would expect the following values of d for each of the wells:

Well 1: $d \approx 0.5$.

Well 2: $d \approx 0.75$.

Well 3: $d \approx 0.25$.

5.7 Estimates of for Logarithmic Acoustic Impedance

Modeling the data using fractionally differenced processes and the regression method outlined in Section 5.2, gives the results shown in Tables 5.5 and 5.6. Table 5.5 shows results obtained by setting $NW = 4$, which allows us to use up to 7 tapers in the multitaper estimates, while Table 5.6 shows the results obtained by setting $NW = 6$, allowing us to use up to 11 tapers. The series all have more than 1200 data points, allowing us to use more tapers than we could for shorter series.

We can see from Table 5.6 that for the first two wells the estimates of d follow a generally increasing pattern as the number of tapers increases – this increasing

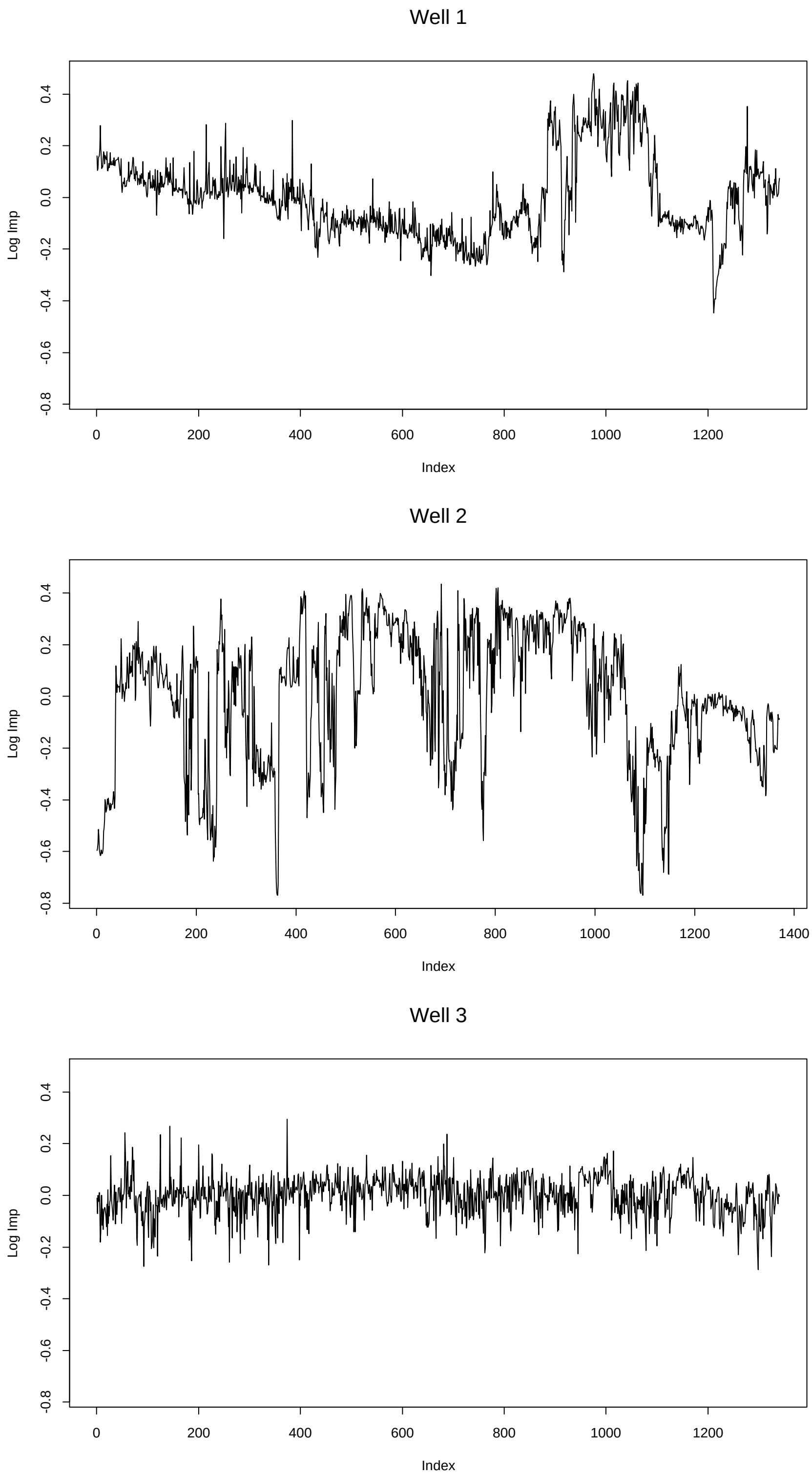


Figure 5.14: Logarithmic impedance

behaviour is in keeping with the simulations results showing the increasing positive bias as the number of tapers used approaches the Shannon number. The third well does not follow this pattern, but first decreases before increasing.

The results appear to agree with the level of dependence found in Walden & Hosken (1985), with the third well being the only one whose value of d lies within the bounds of stationarity.

Figures 5.15 and 5.16 show the resulting multitaper spectral estimates, and we can clearly see that the variance of the estimates reduces as the number of tapers used increases, allowing us to obtain improved estimates of the fractional differencing parameter. As these series are quite long we can afford to use a much larger defining bandwidth, and so are able to increase the number of tapers used in the spectral estimates.

Number of Tapers	Estimates of d for:		
	Well 1	Well 2	Well 3
3	0.6556	0.7462	0.2802
4	0.6490	0.7488	0.2744
5	0.6576	0.7447	0.2799
6	0.6592	0.7460	0.2962
7	0.6673	0.7453	0.3121

Table 5.5: Impedance: Estimates of d using $NW = 4$

Number of Tapers	Estimates of d for:		
	Well 1	Well 2	Well 3
3	0.6619	0.7430	0.2841
4	0.6557	0.7499	0.2831
5	0.6561	0.7466	0.2795
6	0.6651	0.7463	0.2770
7	0.6605	0.7478	0.2833
8	0.6655	0.7448	0.2900
9	0.6752	0.7441	0.2946
10	0.6778	0.7476	0.3050
11	0.6761	0.7502	0.3156

Table 5.6: Impedance: Estimates of d using $NW = 6$

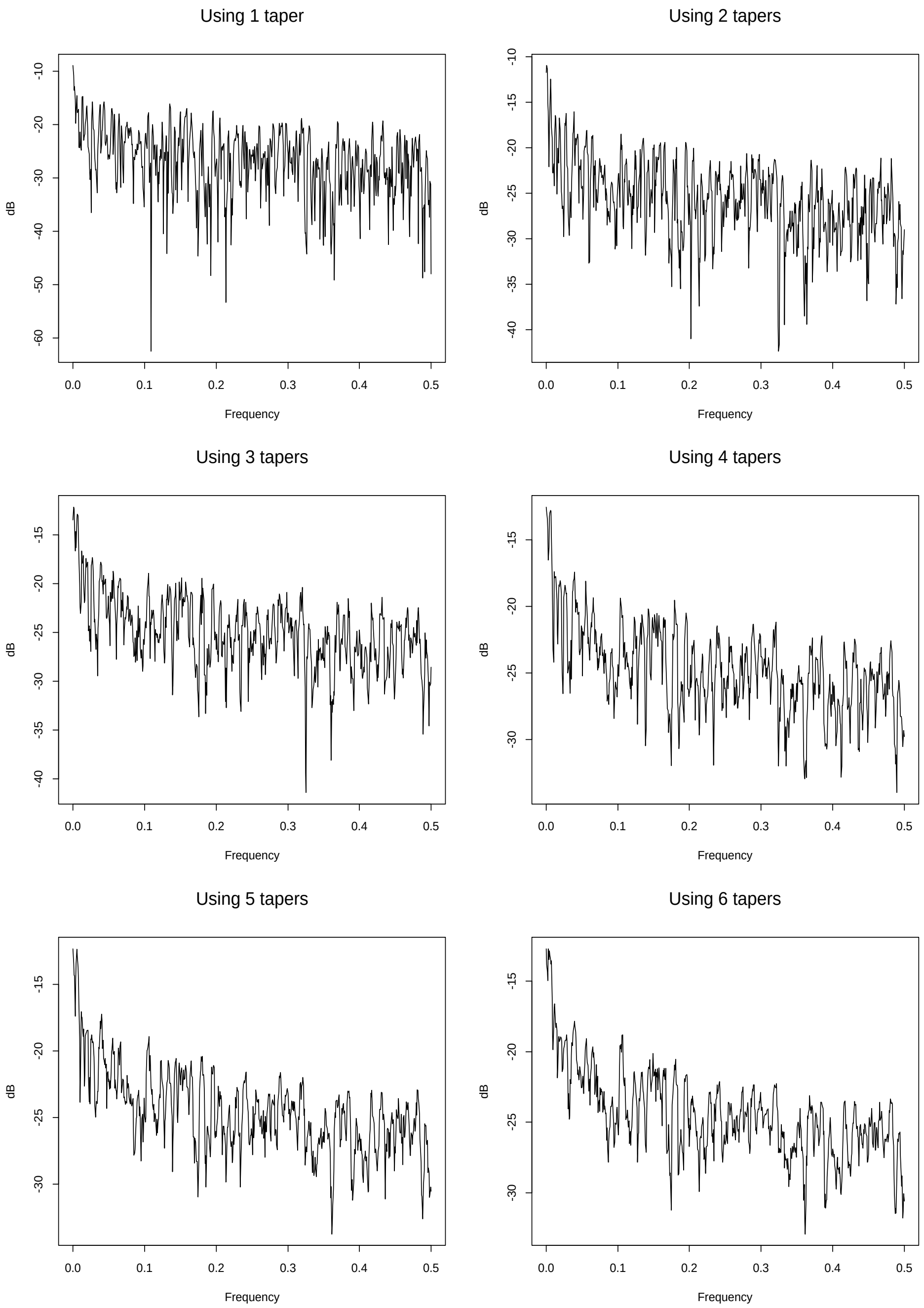


Figure 5.15: Multitaper spectral estimates of the impedance for Well 3, using $NW = 4$

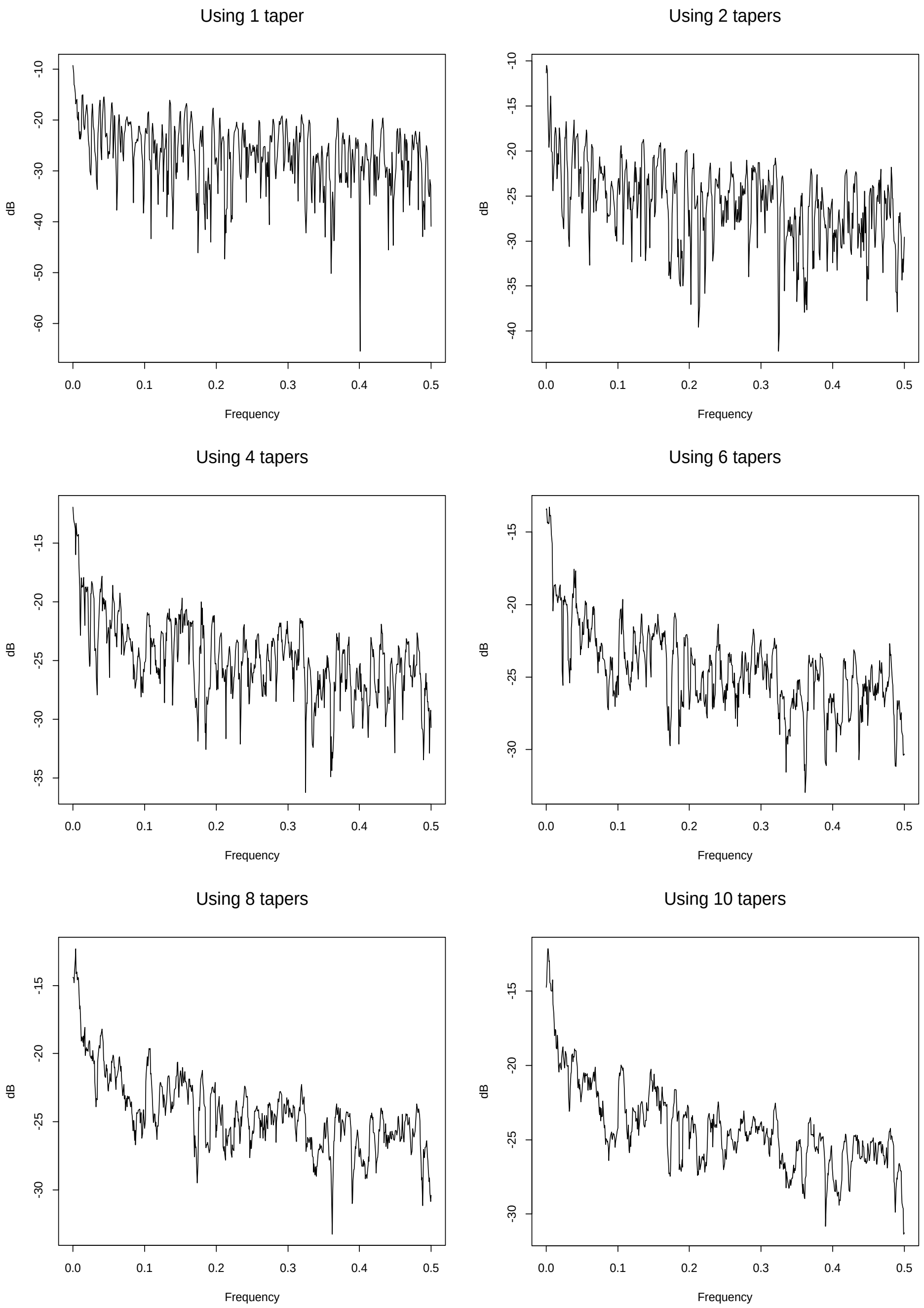


Figure 5.16: Multitaper spectral estimates of the impedance for Well 3, using $NW = 6$

Chapter 6

Wavelets

6.1 Introduction

Wavelet transforms are becoming increasingly popular in all areas of signal and image processing; for examples of some of the current applications of wavelets and an overview of the history of wavelets see Meyer (1993). Wavelets are a relatively recent theory, but despite this, several books have already appeared on the subject, most notably Daubechies (1992a) and Chui (1992). For good introductions to wavelets see Daubechies (1992b), Strang (1989), and Section 13.10 of Press *et al* (1990). For discussions of wavelets from a signal processing viewpoint see Vetterli & Hurley (1992), Rioul & Vetterli (1991) and Vaidyanathan (1993), and for a description in terms of multiresolution theory see Rioul (1993).

In this work we will concentrate on the discrete wavelet transform (DWT), though the interested reader will find a detailed treatment of the continuous wavelet transform in Chui (1992).

From a signal processing point of view, a wavelet is a bandpass filter, and the discrete wavelet transform can be interpreted as a constant-Q filtering with a set of octave band filters, that is, the width of each band as a proportion of its centre frequency is constant (Vetterli and Herley (1992)).

Further, the wavelet analysis of a time series in effect decomposes the variance of the time series into separate components, each one of which can be associated with a particular scale and location. Spectral analysis of a time series, as we have seen earlier, decomposes the variance of the time series into components associated with different frequencies, it is for this reason that a spectral analysis may be unable to effectively analyze time series with discontinuities, or whose structure changes with time. In contrast, the wavelet transform maintains information in the transform about both scale and location.

For an excellent comparison of wavelet transforms and Fourier transforms see Strang (1993).

New work in this chapter include investigating the effect of different boundary conditions on the discrete wavelet transform coefficients, and comparing the phase functions associated with wavelets in current use.

6.2 The wavelet transform

We can think of wavelets as forming an orthonormal basis (see Daubechies (1988) and (1993)), with functions formed by dilating and translating one compact mother wavelet $\psi(\cdot)$. The basis is defined by

$$\psi_n^m(t) = \frac{m}{2}\psi(mt - n).$$

Where m denotes the scale $(\ldots, -8, -4, -2, -1, 0, 1, 2, 4, 8, \ldots)$, and $n \in \mathbf{Z}$ denotes the translation. In this work we have chosen to measure scale directly in powers of two (i.e. $\ldots, -8, -4, -2, -1, 0, 1, 2, 4, 8, \ldots \equiv (-1)^j 2^j,\ j \in \mathbf{N}$) as this directly relates to the physical scale of the time series in question. This is in contrast to the majority of wavelet literature, which has the scale measurement running through the integers (i.e. $m = \ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$), in which case the basis functions are defined by $\psi_n^m(t) = 2^{m/2}\psi(2^m t - n)$.

Using this as an orthonormal basis for all functions in $L^2(\mathbf{R})$, we can define the transform as follows

$$x(t) \leftrightarrow x_n^m.$$

Where we have the usual synthesis/analysis equations for $x(t)$ real

$$\begin{aligned} x(t) &= \textstyle\sum_m \sum_n x_n^m \psi_n^m(t), \\ x_n^m &= \textstyle\int_{-\infty}^{\infty} x(t)\psi_n^m(t)\, dt. \end{aligned}$$

6.2.1 Examples of wavelets

There are many wavelets which satisfy the conditions just described, and which one is employed depends upon the task being undertaken. The unique thing about certain wavelets, is that they can be designed to be bounded in both the frequency and time domains. This is in contrast to the Fourier transform which employs basis functions (complex exponentials) which have perfect resolution in the frequency domain, but are consequently unbounded in the time domain.

Some wavelets, while having excellent time resolution, have poor frequency resolution – the simplest of all wavelets the Haar wavelet, which we shall now describe, is an example of such a wavelet. There are also wavelets with excellent frequency resolution, but consequently poor time resolution. Our aim is to strike a balance and use a wavelet which has reasonable resolution in both the frequency and time domain. Details relating to the construction of wavelets can be found in Cohen & Daubechies (1993) and Strichartz (1993).

The Haar wavelet

The simplest example of a wavelet basis is the Haar basis, which can be generated from the mother wavelet

$$\psi(x) = \begin{cases} +1 & 0 \le x < \frac{1}{2} \\ -1 & \frac{1}{2} \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

While this basis may prove adequate in some applications, as we shall see later, it has very poor frequency resolution. This fact has led to the development of many other wavelets whose mother wavelet has much higher regularity.

Daubechies wavelets

Two families of wavelets which have become very popular are those created by Daubechies (1992a); the *extremal phase* (EP) wavelets and the *least asymmetric* (LA) wavelets. Both sets of wavelets have members with increasing regularity resulting in improved frequency resolution, but consequently with longer time domain support, resulting in poorer time resolution. A wavelet can therefore be selected with the required balance between time and frequency resolution. Details of these wavelets and their construction can be found in Daubechies (1992a).

The Haar wavelet, and some examples of both the extremal phase and least asymmetric wavelets are shown in Figure 6.1, showing the increasing regularity of the wavelets. Every wavelet has an associated filter, and the numbers in the figure associated with the EP and LA wavelets refers to the number of coefficients in these filters. These filters define the discrete wavelet transform, which we shall now describe.

6.3 The discrete wavelet transform

Recently, fast algorithms have been developed for the construction of the discrete wavelet transform, for example see Mallat (1989), Rioul & Duhamel (1992) and Nason & Silverman (1994).
The discrete wavelet transform employs the use of two filters of length L:

- a low-pass filter $\{g_k\}$ for $k = 0, \ldots L - 1$,
- a high-pass filter $\{h_k\}$ for $k = 0, \ldots L - 1$.

If we are interested in an orthogonal wavelet transform then it is the case that

$$g_k = (-1)^{k+1} h_{L-k-1} \quad \text{for } k = 0, \ldots L - 1. \tag{6.1}$$

The discrete wavelet transform starts by simultaneously passing the data through the low and high-pass filters, but only taking every other element of the output (decimation). This results in two vectors, both half as long as the original data set. The decimated output of the high-pass filter gives us our first set of 'detail' or 'wavelet' coefficients. The output from the low-pass filter is used to continue the process, by subsequently passing that though the low and high-pass filters. The algorithm is outlined in greater detail in Table 6.1.

Given low and high-pass filters with special properties ensures that the transform is orthogonal, and then an inverse wavelet transform follows in a very similar

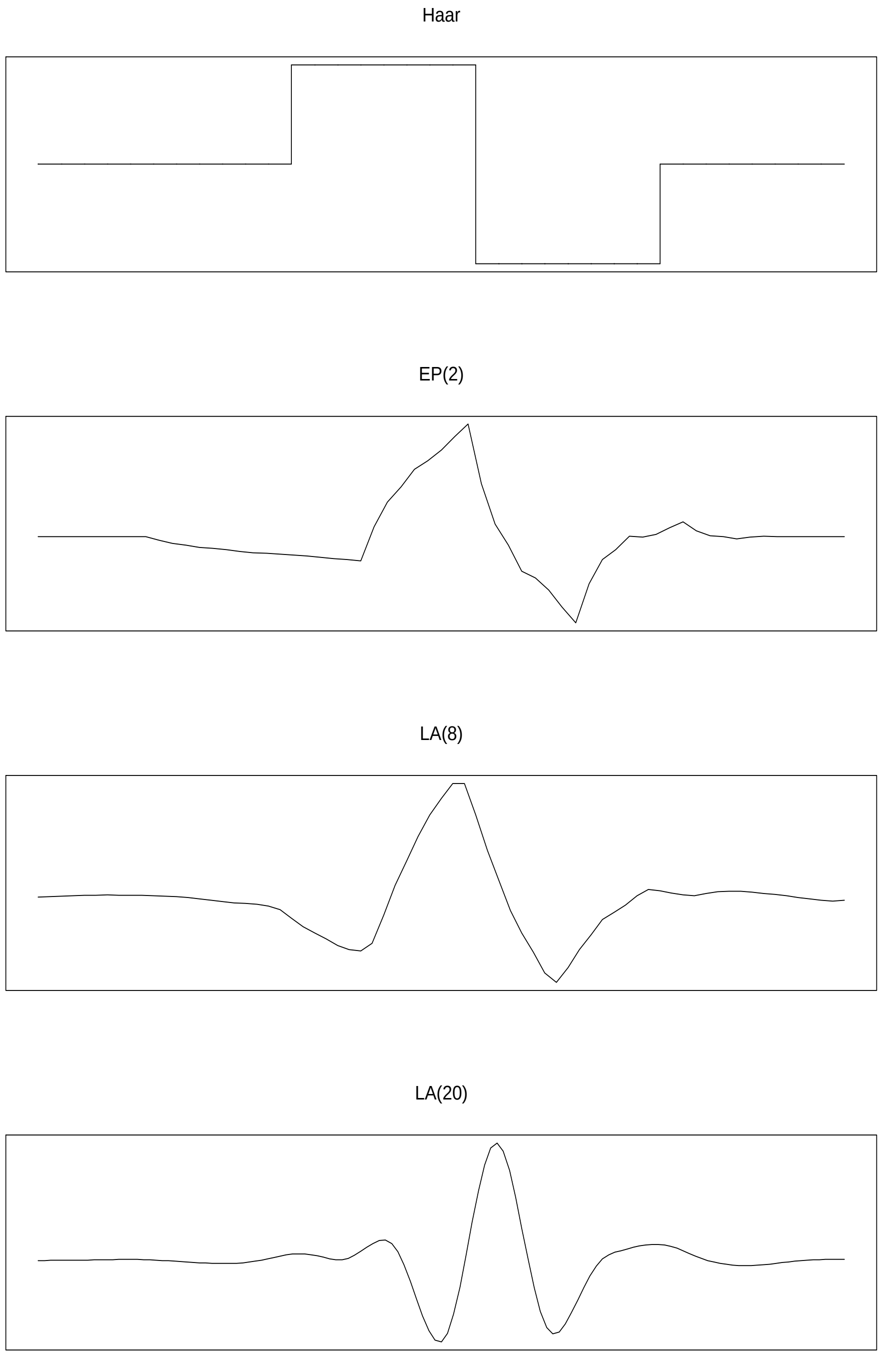


Figure 6.1: Sampling of wavelets

1. Given a time series

$$x_1^0, \ldots, x_N^0,$$

where the superscript denotes scale.

2. Form two series of length $N/2$:

$$x_1^1, \ldots, x_{N/2}^1 \quad \text{and} \quad d_1^1, \ldots, d_{N/2}^1.$$

The $\{x_j^1\},\ j = 1, \ldots, N/2$ is the result of low-pass filtering $\{x_j^0\}$ with $\{g_l\}$ and taking every other point in the output (decimated).
The $\{d_j^1\},\ j = 1, \ldots, N/2$ is the result of high-pass filtering $\{x_j^0\}$ with $\{h_l\}$ and taking every second point of the output (decimated).

3. Then at subsequent scales m form two series of length $N/2m$

$$x_1^m, \ldots, x_{N/2m}^m \quad \text{and} \quad d_1^m, \ldots, d_{N/2m}^m.$$

The $\{x_j^m\}$ is the decimated output of low-pass filtering $\{x_j^{m/2}\}$.
The $\{d_j^m\}$ is the decimated output of high-pass filtering $\{x_j^{m/2}\}$
(See Figure 6.2)

4. Repeat step 3 until the transform vector is obtained

$$\left(x_1^{N/2}, d_1^{N/2}, d_1^{N/4}, d_2^{N/4}, \ldots, d_1^1, \ldots, d_{N/2}^1\right)^T$$

containing $N-1$ 'detail' coefficients –

$$d_{j,m}, \qquad m = 1, 2, 4, 8, \ldots, N/2 \text{ and } j = 1, 2, 3, \ldots, N/2m,$$

and one 'smooth' coefficient – $x_1^{N/2}$.

Table 6.1: Algorithm describing the discrete wavelet transform

way to the forward transform. A step in the inverse transform is described in Figure 6.3. Here H and G represent the Fourier transforms of the filters $\{h_l\}$ and $\{g_l\}$. These filters, when applied in this iterative fashion, converge to the wavelet functions.

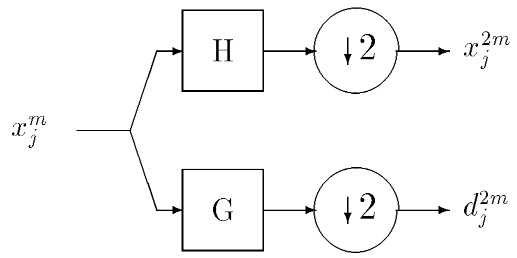


Figure 6.2: Discrete Wavelet Transform: Decomposition step

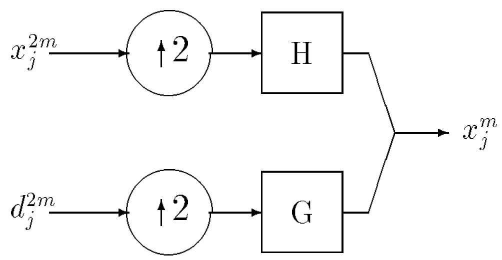


Figure 6.3: Discrete Wavelet Transform: Reconstruction step

Figure 6.4 shows the discrete wavelet transform in action on a particular dataset, displaying the output of the filters at successive scales. The actual filters applied at different scales are shown in Figures 6.5 to 6.8 for the wavelets that were illustrated in Figure 6.1. These figures show the iterated filters and also their transfer functions at increasing scales, clearly demonstrating the bandpass filter nature of the discrete wavelet transform, and showing the compromise that must be made between time and frequency resolution.

6.3.1 Matrix representation

If we are using a low-pass filter whose coefficients are $\{g_0, g_1, g_2, g_3\}$, then due to the relationship described in Equation 6.1, the corresponding high-pass filter

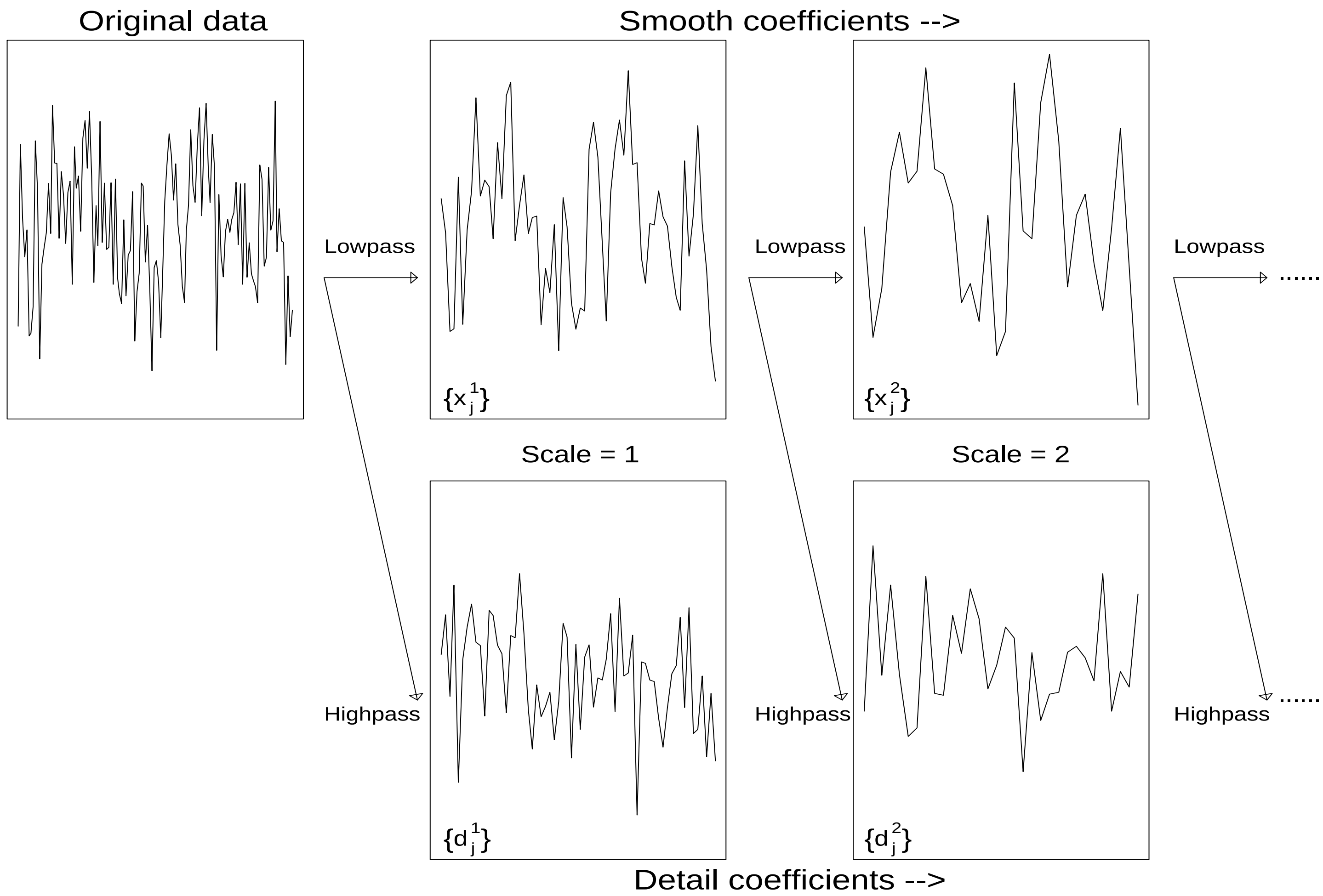


Figure 6.4: Demonstration of the discrete wavelet transform

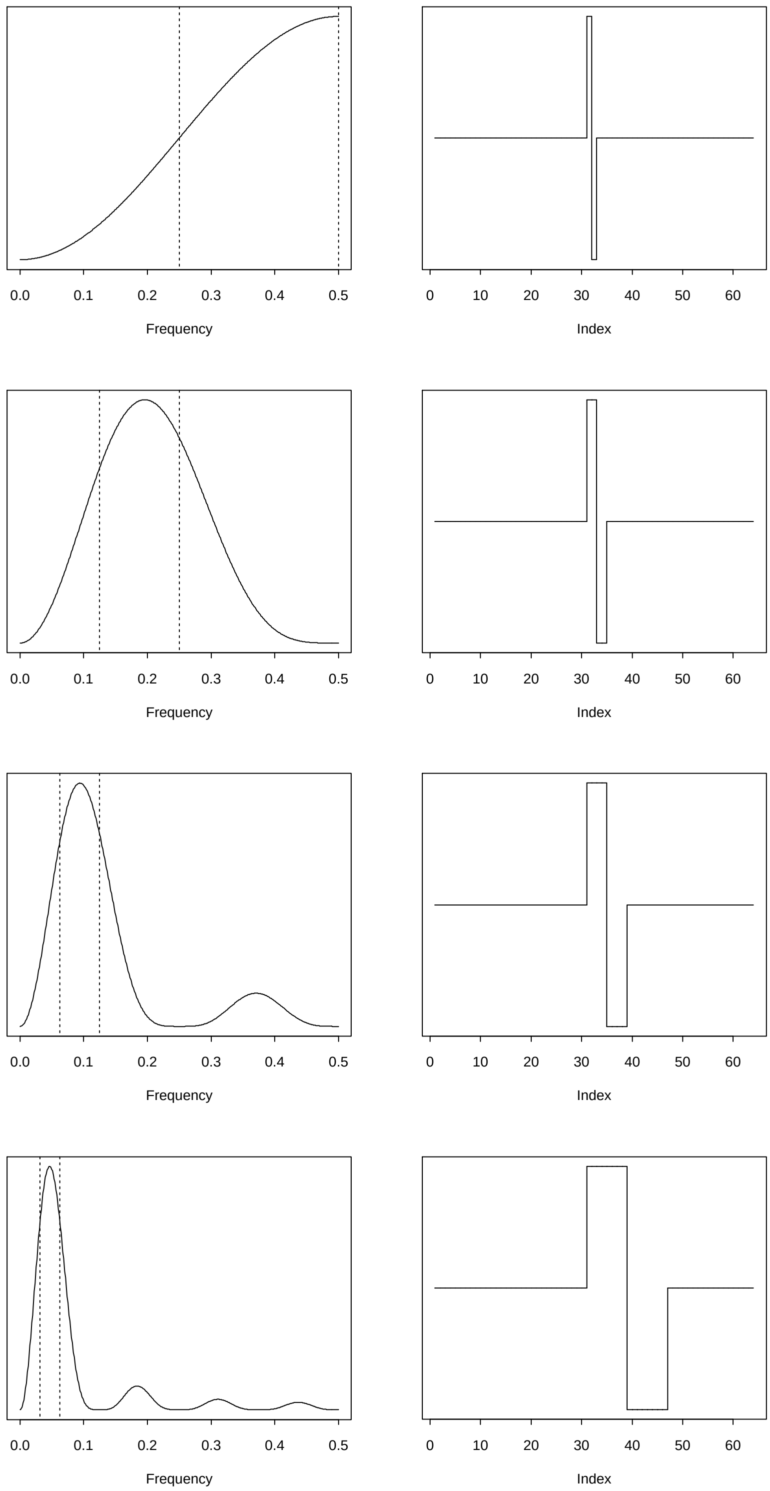


Figure 6.5: Haar: Iterated filters and transfer functions

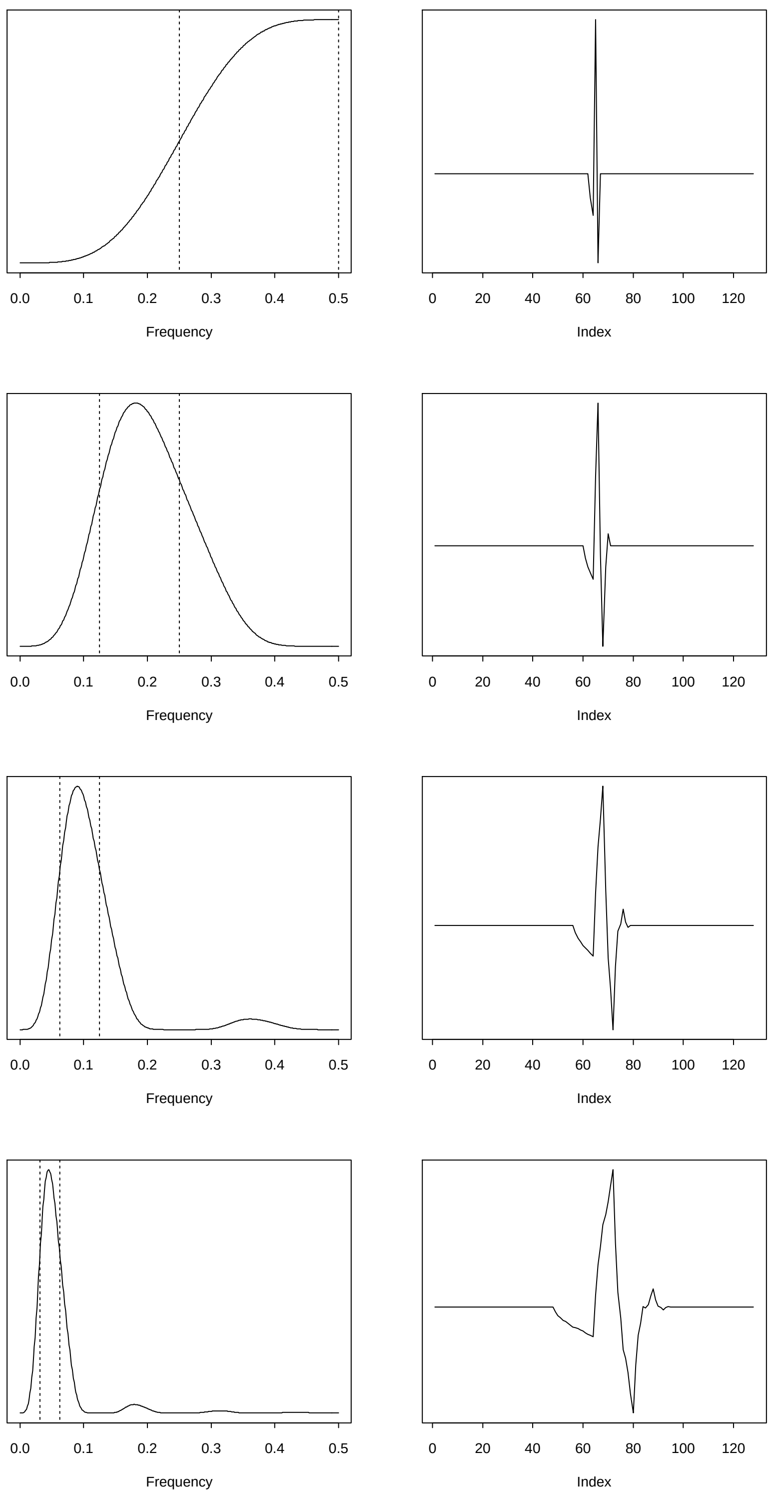


Figure 6.6: EP(4): Iterated filters and transfer functions

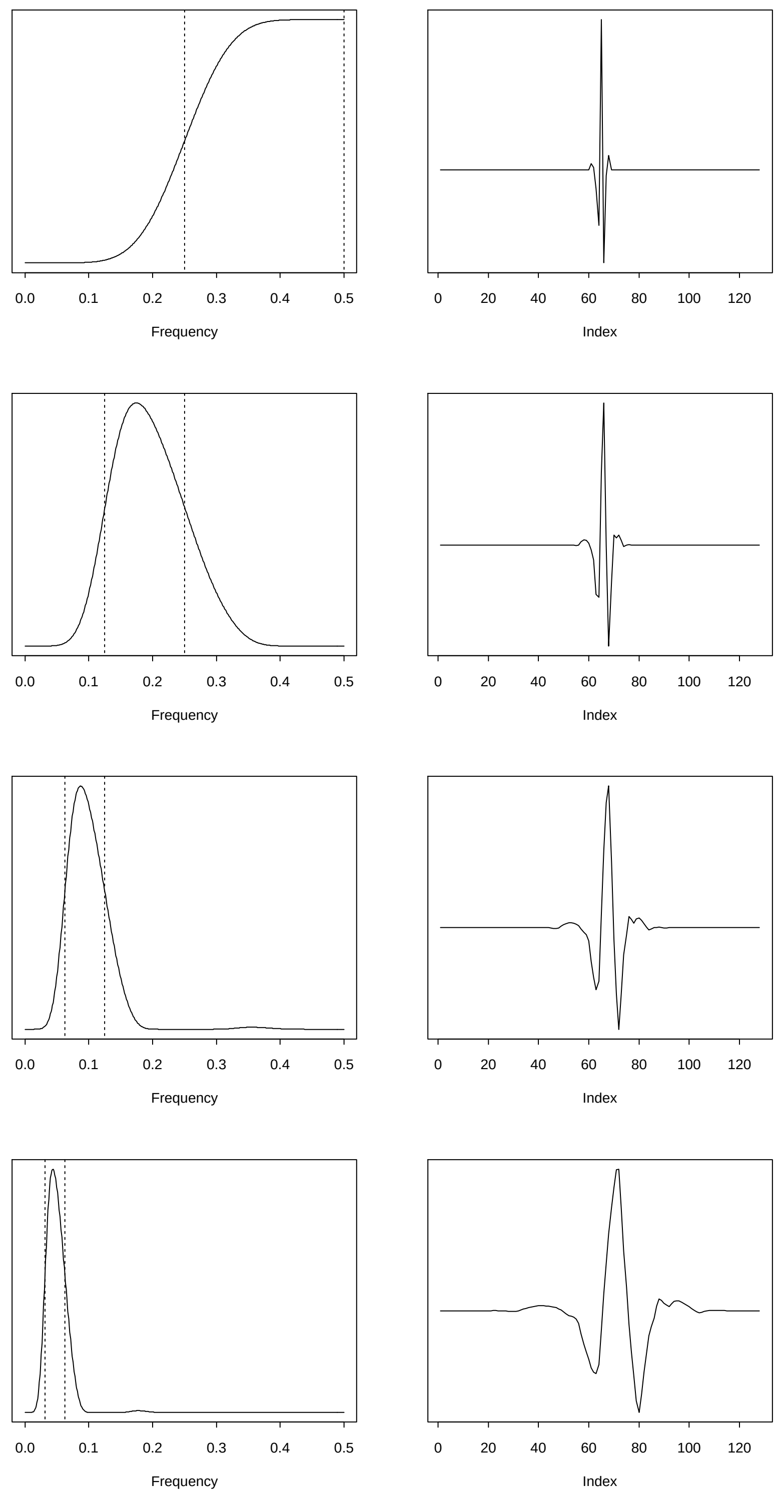


Figure 6.7: LA(8): Iterated filters and transfer functions

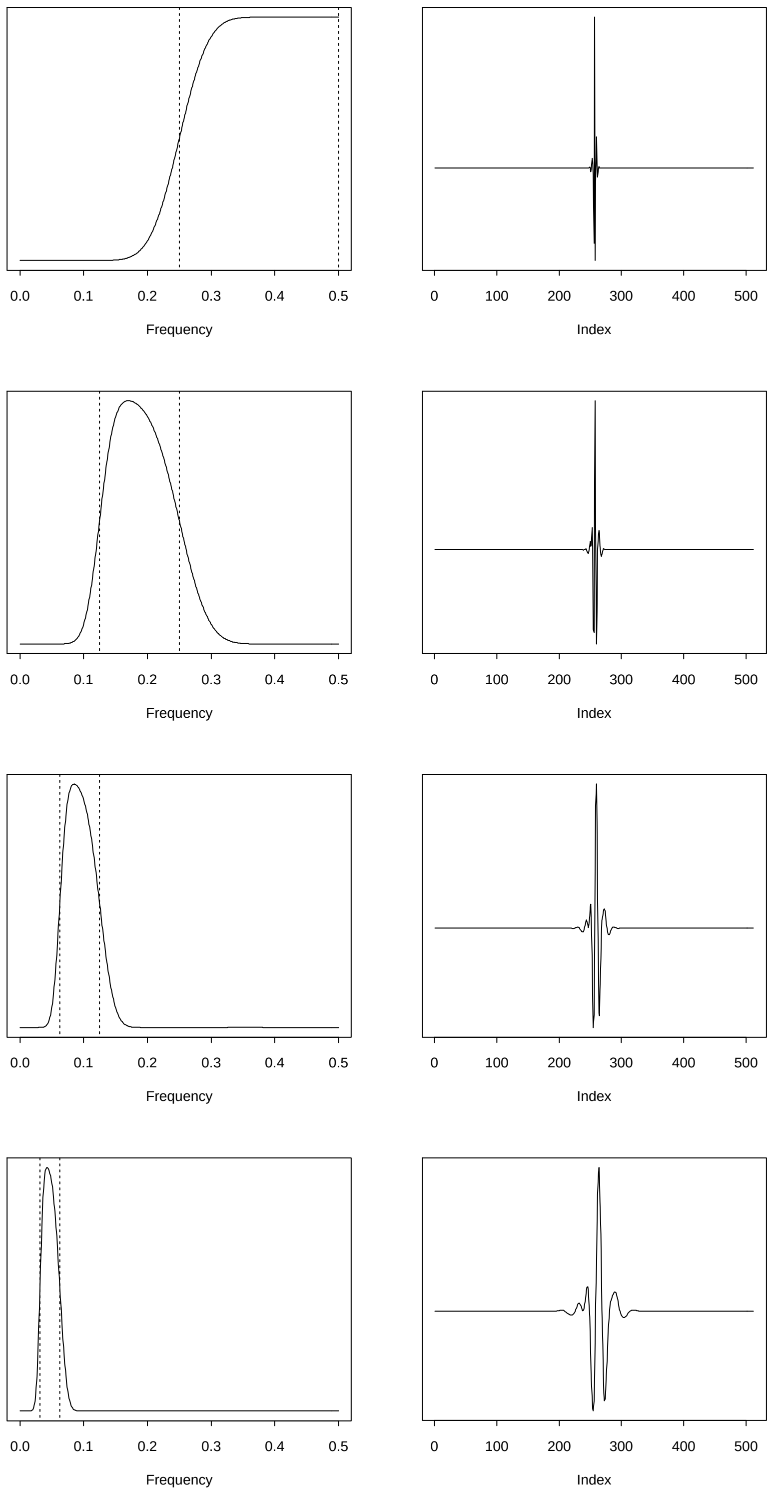


Figure 6.8: LA(20): Iterated filters and transfer functions

coefficients are $\{g_3, -g_2, g_1, -g_0\}$. Each step of the discrete wavelet transform can be expressed in terms of the matrix equation:

$$T^m X^m = Y^{2m} \qquad m = 1, 2, 4, 8, \ldots, \frac{N}{4}.$$

The matrix T^m, describing the transform at scale m is of dimension $(N/2m, N/2m)$, and the vectors X^m and Y^{2m} for $m \geq 1$, are given by:

$$\begin{bmatrix} g_0 & g_1 & g_2 & g_3 & & & & & & & \\ g_3 & -g_2 & g_1 & -g_0 & & & & & & & \\ & & g_0 & g_1 & g_2 & g_3 & & & & & \\ & & g_3 & -g_2 & g_1 & g_0 & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & g_0 & g_1 & g_2 & g_3 \\ & & & & & & & g_3 & -g_2 & g_1 & g_0 \\ g_2 & g_3 & & & & & & & & g_0 & g_1 \\ g_1 & -g_0 & & & & & & & & g_3 & -g_2 \end{bmatrix} \begin{bmatrix} x_1^m \\ x_2^m \\ x_3^m \\ x_4^m \\ \vdots \\ x_{N/2m-3}^m \\ x_{N/2m-2}^m \\ x_{N/2m-1}^m \\ x_{N/2m}^m \end{bmatrix} = \begin{bmatrix} x_1^{2m} \\ d_1^{2m} \\ x_2^{2m} \\ d_2^{2m} \\ \vdots \\ x_{N/4m-1}^{2m} \\ d_{N/4m-1}^{2m} \\ x_{N/4m}^{2m} \\ d_{N/4m}^{2m} \end{bmatrix}. \tag{6.2}$$

Any filters of length greater than two will automatically produce boundary problems. Here we have assumed that the data is periodic, which results in the coefficients of the filter 'wrapping' around in the matrix T shown above. These boundary conditions are implemented at every stage in the discrete wavelet transform, and we shall investigate potential problems with these imposed boundary conditions in Section 6.5.

6.4 Matrix form of the discrete wavelet transform

The discrete wavelet transform is an orthogonal transform of the data, and as such, can be described in term of a matrix transformation. Given a time series

$$X_N = \{x_1, \ldots, x_N\}$$

and its discrete wavelet transform

$$Z = \{x_1^{N/2}, d_1^{N/2}, d_1^{N/4}, d_2^{N/4}, \ldots, d_1^1, \ldots, d_{N/2}^1\}.$$

Then the following relationship holds for the wavelet matrix W:

$$Z = WX_N.$$

We will now proceed to establish the exact form of W.

6.4.1 High and low-pass matrices

We can split the matrix T^m given in Equation 6.2 into two matrices: R^m describing the low-pass filtering, and Q^m describing the high-pass filtering. For a wavelet whose associated filter has four components these matrices have the following form:

$$R^m = \begin{bmatrix} g_0 & g_1 & g_2 & g_3 & & & & & & & & \\ & & g_0 & g_1 & g_2 & g_3 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & & g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & & & & & & & & & g_0 & g_1 \end{bmatrix}, \qquad (6.3)$$

$$Q^m = \begin{bmatrix} g_3 & -g_2 & g_1 & -g_0 & & & & & & & & \\ & & g_3 & -g_2 & g_1 & g_0 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & & g_3 & -g_2 & g_1 & g_0 \\ g_1 & -g_0 & & & & & & & & & g_3 & -g_2 \end{bmatrix}.$$

Notice that each of these matrices have half as many rows as columns, and at scale m, will have dimension $(N/4m \times N/2m)$. Now we can describe the discrete wavelet transform in terms of operations using these two matrices.

We now have the following relationship between the 'smooth' coefficients as we move from scale m to scale $2m$:

$$R^m X^m = X^{2m} \quad \text{for} \quad m = 1, \ldots, N/4,$$

where X^{2m} is defined as the vector of 'smooth' coefficients $\{x_1^{2m}, x_2^{2m}, \ldots, x_{N/4m}^{2m}\}$. Similarly, we can establish the following relationship between the 'smooth' coefficients at scale m, and the 'detail' coefficients at the next scale $2m$:

$$Q^m X^m = D^{2m} \quad \text{for} \quad m = 1, \ldots, N/4,$$

where D^{2m} is defined as the vector of 'detail' coefficients $\{d_1^{2m}, d_2^{2m}, \ldots, d_{N/4m}^{2m}\}$. Let Q_{jk}^m be the (j,k)th element of Q^m, then

$$d_k^{2m} = \sum_{j=1}^{N/2m} x_j^m Q_{kj}^m \quad \text{for} \quad \begin{cases} m &= 1,\ldots,N/4 \\ k &= 1,\ldots,N/4m. \end{cases} \tag{6.4}$$

Let R_{jk}^m be the (j,k)th element of R^m, then

$$x_k^{2m} = \sum_{j=1}^{N/2m} x_j^m R_{kj}^m \quad \text{for} \quad \begin{cases} m &= 1,\ldots,N/4 \\ k &= 1,\ldots,N/4m. \end{cases} \tag{6.5}$$

For the first step in the transform we have

$$x_k^1 = \sum_{j=1}^{N} x_j R_{kj}^0 \ , \quad d_k^1 = \sum_{j=1}^{N} x_j Q_{kj}^0 \quad \text{for} \quad k = 1,\ldots,N/2.$$

From Equation 6.5 we have

$$\begin{aligned} x_k^2 &= \textstyle\sum_{j=1}^{N/2} x_j^1 R_{kj}^1 \\ &= \textstyle\sum_{j=1}^{N/2} \sum_{l=1}^{N} x_l R_{jl}^0 R_{kj}^1 \\ &= \textstyle\sum_{l=1}^{N} x_l (R^1 R^0)_{kl} \quad \text{for} \quad k = 1,\ldots,N/4. \end{aligned}$$

Continuing in this fashion we see that at scale m,

$$x_k^{2m} = \sum_{t=1}^{N} x_t (R^m R^{m/2} \ldots R^0)_{kt} \quad \text{for} \quad \begin{cases} m &= 1,\ldots,N/4 \\ k &= 1,\ldots,N/4m. \end{cases} \tag{6.6}$$

Now combining this with Equation 6.4 gives

$$d_k^{2m} = \sum_{t=1}^{N/2m} Q_{kt}^m \left[\sum_{j=1}^{N} x_j (R^{m/2} R^{m/4} \ldots R^0)_{tj} \right].$$

Hence

$$d_k^{2m} = \sum_{t=1}^{N} x_t (Q^m R^{m/2} R^{m/4} \ldots R^0)_{kt} \quad \text{for} \quad \begin{cases} m &= 1,\ldots,N/4 \\ k &= 1,\ldots,N/4m. \end{cases}$$

Now we have general definitions of the 'detail' coefficients, we can construct the form of W

$$\begin{array}{c} \frac{N}{2} : (1 \times N) \\ \frac{N}{2} : (1 \times N) \\ \frac{N}{4} : (2 \times N) \\ \vdots \\ 2 : (N/4 \times N) \\ 1 : (N/2 \times N) \end{array} \begin{pmatrix} R^{N/4} \ldots R^0 \\ Q^{N/4} R^{N/8} \ldots R^0 \\ Q^{N/8} R^{N/16} \ldots R^0 \\ \vdots \\ Q^1 R^0 \\ Q^0 \end{pmatrix} \tag{6.7}$$

where the left-hand border gives the scale and dimensions of each of the blocks that go to make up the matrix.

This orthogonal matrix is illustrated in Figure 6.9, for $N = 128$ using the EP(4) wavelet and periodic boundary conditions. The top diagram of Figure 6.9 shows the first row of the matrix, and involves only smoothing matrices. Each of the subsequent rows in Figure 6.9 display a selected row of the subsequent blocks of the matrix as outlined in Equation 6.7. For example, the third diagram from the top shows the second row of the $(2 \times N)$ matrix $Q^{16}R^8R^4R^2R^1R^0$. The labels represent the element of the discrete wavelet transform that would result from multiplying that particular row by the original data.

6.5 Boundary effects

As we have seen from the description of the discrete wavelet transform, if we are employing a filter of length greater than two, we will be faced with boundary problems. In a time series context, the most appropriate course of action is to assume that the data are periodic. This section will investigate further the effects that this assumption has on the output of the discrete wavelet transform.

6.5.1 Relationship between 'smooth' coefficients and the mean

Let R^m_{jk} be the (j,k)th element of R^m as defined in Section 6.4.1. Then summing over both sides of Equation 6.6 gives

$$\sum_{k=1}^{N/4m} x_k^{2m} = \sum_{t=1}^{N} x_t \sum_{k=1}^{N/4m} (R^m R^{m/2} \dots R^0)_{kt}. \tag{6.8}$$

For the case of periodic boundary conditions with smoothing matrix 6.3, we can readily compute the sum in Equation 6.8 for any scale m. To do this, we use two results on orthogonal wavelet designs; that the general discrete wavelet transform should be invertible requires that

$$\frac{1}{2}\left[|G(f)|^2 + \left|G\left(f+\frac{1}{2}\right)\right|^2\right] = 1,$$

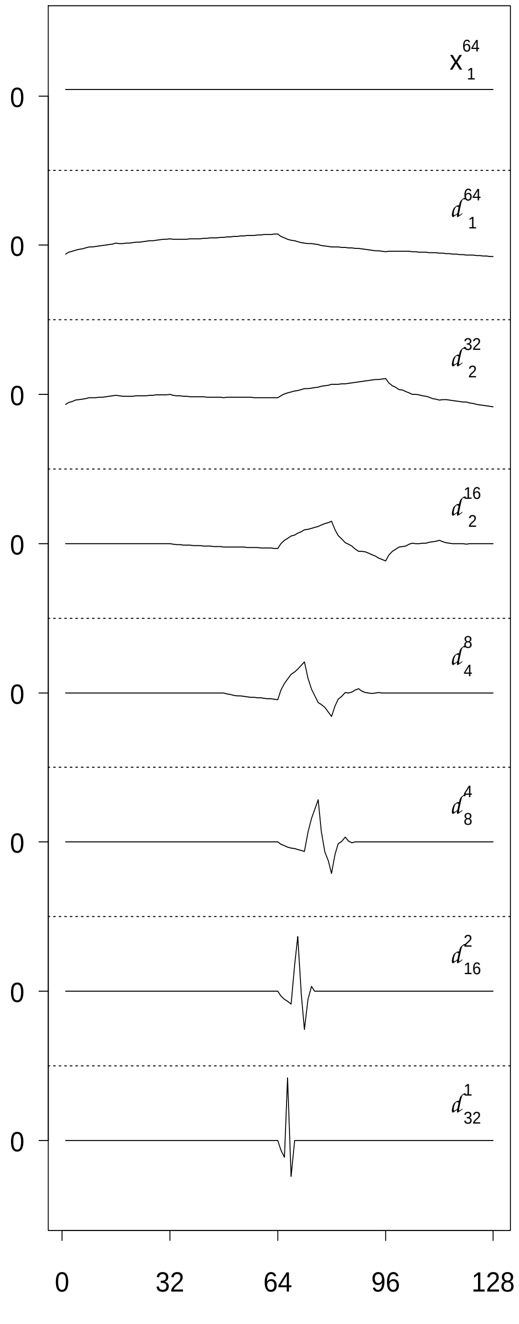


Figure 6.9: Wavelet matrix W, for the EP(4) wavelet

where $G(f) = \sum_l g_l e^{-i2\pi fl}$ is the Fourier transform of the smoothing filter. This is called 'condition O' (orthogonality) in Strang (1989, p.620). This constraint is identical to $\sum g_l g_{l+2k} = \delta_{0k}$, where $k = 1, 2, \ldots$ and δ_{0k} is the Kronecker delta. For the EP(4) coefficients this means that $g_0^2 + g_1^2 + g_2^2 + g_3^2 = 1$ and $g_0 g_2 + g_1 g_3 = 0$. The second condition given in Strang (1989, p.617) called 'condition A' ('approximation' or smoothness) is that $G(f)$ must have a zero of at least first order at $f = 1/2$. The condition $G(1/2) = 0$ is identical to $\sum_l (-1)^l g_l = 0$. For the EP(4) coefficients this gives the result $g_0 + g_2 = g_1 + g_3$.

Summing both sides of Equation 6.5:

$$\sum_{k=1}^{N/4m} x_k^{2m} = \sum_{j=1}^{N/2m} x_j^m \sum_{k=1}^{N/4m} R_{kj}^m. \tag{6.9}$$

We can see from the matrix 6.3 that

$$\sum_{k=1}^{N/4m} R_{kj}^m = g_0 + g_2 \quad \text{or} \quad \sum_{k=1}^{N/4m} R_{kj}^m = g_1 + g_3,$$

as it is always a column sum of the matrix 6.3. But, from Strang's 'condition A' we have seen that $g_0 + g_2 = g_1 + g_3$. Combining this with the fact that $(g_0 + g_2)^2 + (g_3 + g_1)^2 = g_0^2 + g_1^2 + g_2^2 + g_3^2 + 2g_0 g_2 + 2g_1 g_3 = 1$, by using the 'condition O' orthogonality results, it follows that

$$\sum_{k=1}^{N/4m} R_{kj}^m = 1/\sqrt{2} \qquad \forall\, j, m.$$

Hence, by repeated application of Equation 6.9 we obtain

$$\sum_{k=1}^{N/2m} x_k^m = (\sqrt{2})^{-(\log_2 m+1)} \sum_{t=1}^{N} x_t = N(\sqrt{2})^{-(\log_2 m+1)} \bar{x}. \tag{6.10}$$

The values for $g_0 + g_2$ and $g_1 + g_3$ could have been obtained directly from the values of the EP(4) filter, but note that the relationship established in Equation 6.10 holds not only for EP(4) but for any orthogonal wavelet design satisfying the two conditions O and A described above.

Most importantly, from Equation 6.10, we can calculate the relationship between the only 'smooth' coefficient left in the discrete wavelet transform – $x_1^{N/2}$ and the original data. The importance of this relationship will become clear in Chapter 7, where we shall employ the use of the inverse discrete wavelet transform to synthesise long memory processes.

6.5.2 Wavelets defined on $[0,1]$

In an attempt to counteract possible problems associated with edge effects, Cohen *et al* (1993a,b) look at the case of orthonormal wavelet bases for functions defined on $[0,1]$ rather than the whole real line. The algorithm proceeds in an almost identical method to the standard discrete wavelet transform, except that coefficients that are situated at a distance less than $L/2$ points from the edges are found using special filters that correspond to the scaling relations satisfied by the scaling functions and the wavelets constructed at the edges. For example, if using the EP(4) wavelet filter coefficients then we require $L/2 = 2$ left and 2 right low-pass and high-pass filters; the filters are of length 3 closest to the edge, and length 5 closest to the interior. The $N/2m \times N/2m$ transformation matrix T^m introduced in Equation 6.2 is now modified to

$$T^m = \left[\begin{array}{ccccccccccccccc}
\alpha_0^L & \alpha_1^L & \alpha_2^L & & & & & & & & & & & & \\
\beta_0^L & \beta_1^L & \beta_2^L & & & & & & & & & & & & \\
\gamma_0^L & \gamma_1^L & \gamma_2^L & \gamma_3^L & \gamma_4^L & & & & & & & & & & \\
\delta_0^L & \delta_1^L & \delta_2^L & \delta_3^L & \delta_4^L & & & & & & & & & & \\
 & & & g_0 & g_1 & g_2 & g_3 & & & & & & & & \\
 & & & g_3 & -g_2 & g_1 & -g_0 & & & & & & & & \\
 & & & & & & & \ddots & & & & & & & \\
 & & & & & & & & g_0 & g_1 & g_2 & g_3 & & & \\
 & & & & & & & & g_3 & -g_2 & g_1 & -g_0 & & & \\
 & & & & & & & & & & \gamma_0^R & \gamma_1^R & \gamma_2^R & \gamma_3^R & \gamma_4^R \\
 & & & & & & & & & & \delta_0^R & \delta_1^R & \delta_2^R & \delta_3^R & \delta_4^R \\
 & & & & & & & & & & & & \alpha_0^R & \alpha_1^R & \alpha_2^R \\
 & & & & & & & & & & & & \beta_0^R & \beta_1^R & \beta_2^R
\end{array}\right]$$

Here $\{\alpha_k^L\}$ and $\{\gamma_k^L\}$ are low-pass left-edge filters, $\{\beta_k^L\}$ and $\{\delta_k^L\}$ are high-pass left-edge filters. Note that the smallest number of coefficients that this matrix can act on is 8, so the algorithm stops at scale $N/8$, at which point four 'smooth' and four 'detail' coefficients are computed, another wavelet could be chosen to compute the final coefficients at higher scales if required. The matrix T^m is designed to be orthogonal so that its inverse is equal to its transpose.

The $N/2m \times N/2m$ smoothing matrix R^m for this end-corrected transform takes the form

$$R^m = \begin{bmatrix} \alpha_0^L & \alpha_1^L & \alpha_2^L & & & & & & & & & & \\ \gamma_0^L & \gamma_1^L & \gamma_2^L & \gamma_3^L & \gamma_4^L & & & & & & & & \\ & & & g_0 & g_1 & g_2 & g_3 & & & & & & \\ & & & & & & & \ddots & & & & & \\ & & & & & & & & g_0 & g_1 & g_2 & g_3 & & \\ & & & & & & & & & & \gamma_0^R & \gamma_1^R & \gamma_2^R & \gamma_3^R & \gamma_4^R \\ & & & & & & & & & & & & \alpha_0^R & \alpha_1^R & \alpha_2^R \end{bmatrix}.$$

Again using Equation 6.8 we can readily evaluate the weights attaching to the original data $\{x_1, \ldots, x_n\}$ that define the sum of the 'smooth' coefficients.

On the infinite real line the sequence $\{\ldots, 1, 1, 1, 1, 1, \ldots\}$ is invariant under low-pass filtering and decimation since the wavelets have R (typically several) vanishing moments. On the interval, despite the fact that the wavelets have vanishing moments, the sequence $\{1, 1, 1, \ldots, 1, 1, 1\}$ is not mapped by the low-pass filtering plus decimation to a sequence of 1's of half the original length. The reason for this is that the low-pass filters have been adapted. To circumvent this disadvantage, Cohen *et al* (1993a,b) suggest a data preconditioning step. This step essentially changes $\{1, 1, 1, \ldots, 1, 1, 1\}$ into another sequence that maps to an all zero sequence under high-pass decimation. To accommodate this preconditioning step let us now define $\{x_1, \ldots, x_n\} = \{x_1^{-1}, \ldots, x_n^{-1}\}$. Then the preconditioning step is a matrix operation which changes $\{x_1^{-1}, \ldots, x_n^{-1}\}$ to a new starting vector $\{x_1^0, \ldots, x_1^0\}$ using a $N \times N$ matrix P:

$$PX^{-1} = X^0.$$

For $L = 4$ (that is when using the EP(4) wavelets), this matrix takes the form

$$P = \begin{bmatrix} a & b & & & & & & & \\ 0 & c & & & & & & & \\ & & 1 & & & & & & \\ & & & 1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & d & 0 \\ & & & & & & & e & f \end{bmatrix}.$$

The upper and lower triangular preconditioning matrices consist of non-zero elements a, b, c and d, e, f respectively. For general L, the upper and lower triangular preconditioning matrices are $L/2 \times L/2$.

Let P_{jk} be the (j, k)th element of P. Then

$$x_j^{-1} = \sum_{k=1}^{N} P_{jk} x_k^{p+1}.$$

So,

$$\sum_{k=1}^{N/2m} x_k^m = \sum_{t=1}^{N} \left[\sum_{l=1}^{N} P_{tl} x_l \right] \sum_{k=1}^{N/2m} (R^{m/2} \ldots R^0)_{kt} = \sum_{l=1}^{N} x_l \sum_{k=1}^{N/2m} (R^{m/2} \ldots R^0 P)_{kl},$$

and again we can readily calculate the weights attaching to each of the original data $x_1, \ldots, x_N$, that form the sum of the 'smooth' coefficients.

Figure 6.10 shows the weights for the three methods of treating boundaries discussed here; periodic, unpreconditioned and preconditioned, for $N = 128$. The plots display the weights as the algorithm progresses through the scales. For the plot showing the weights for periodic boundary conditions the solid line represents the weights attached to the original data for the sum of the smooths at the highest scale, which is $x_1^{N/2}$. As we have already mentioned , the other methods of boundary correction looked at here cannot be extended to the final scale, but due to the nature of the filters they employ, the algorithm finishes at scale $N/8$.

In Chapter 7, we develop a method of simulating long memory processes by means of the inverse discrete wavelet transform, and as such will need to provide elements for the vector to be inverse transformed. Here we have shown that as long as we employ periodic boundary conditions we are guaranteed that the sample mean of the resulting data will be directly related to the final smooth coefficient – $x_1^{N/2}$, as opposed to some strange linear combination of the original data.

6.6 Correlation of transform coefficients

Recall from Section 6.4 that the discrete wavelet transform can be written in terms of a matrix equation

$$Z = WX_N.$$

We can immediately derive the variance and expected value of Z as

$$E\{Z\} = WE\{X_N\} \quad \text{and} \quad \text{var}\{Z\} = W\text{var}\{X_N\}W^T.$$

Given a stationary process for which we know the autocovariance sequence, we can easily calculate the variance of the transform coefficients

$$\text{Var}\{Z\} = \begin{pmatrix} w_{1,1} & \dots & w_{1,N} \\ w_{2,1} & \dots & w_{2,N} \\ \vdots & \ddots & \vdots \\ w_{N,1} & \dots & w_{N,N} \end{pmatrix} \begin{pmatrix} s_0 & s_1 & \dots & s_{N-1} \\ s_1 & s_0 & \dots & s_{N-2} \\ \vdots & & \ddots & \vdots \\ s_{N-1} & s_{N-2} & \dots & s_0 \end{pmatrix} \begin{pmatrix} w_{1,1} & \dots & w_{N,1} \\ w_{1,2} & \dots & w_{N,2} \\ \vdots & \ddots & \vdots \\ w_{1,N} & \dots & w_{N,N} \end{pmatrix}. \tag{6.11}$$

6.6.1 Fractionally differenced processes

Recall from Section 2.2 that the autocovariance sequence of a fractionally differenced process with $\sigma_\epsilon^2 = 1$ is given by

$$s_\tau = E(x_t x_{t-\tau}) = \frac{(-1)^\tau(-2d)!}{(\tau - d)!(-\tau - d)!}. \tag{6.12}$$

Substituting this into Equation 6.11 for various forms of W enables us to investigate the exact correlation structure of the discrete wavelet transform of a

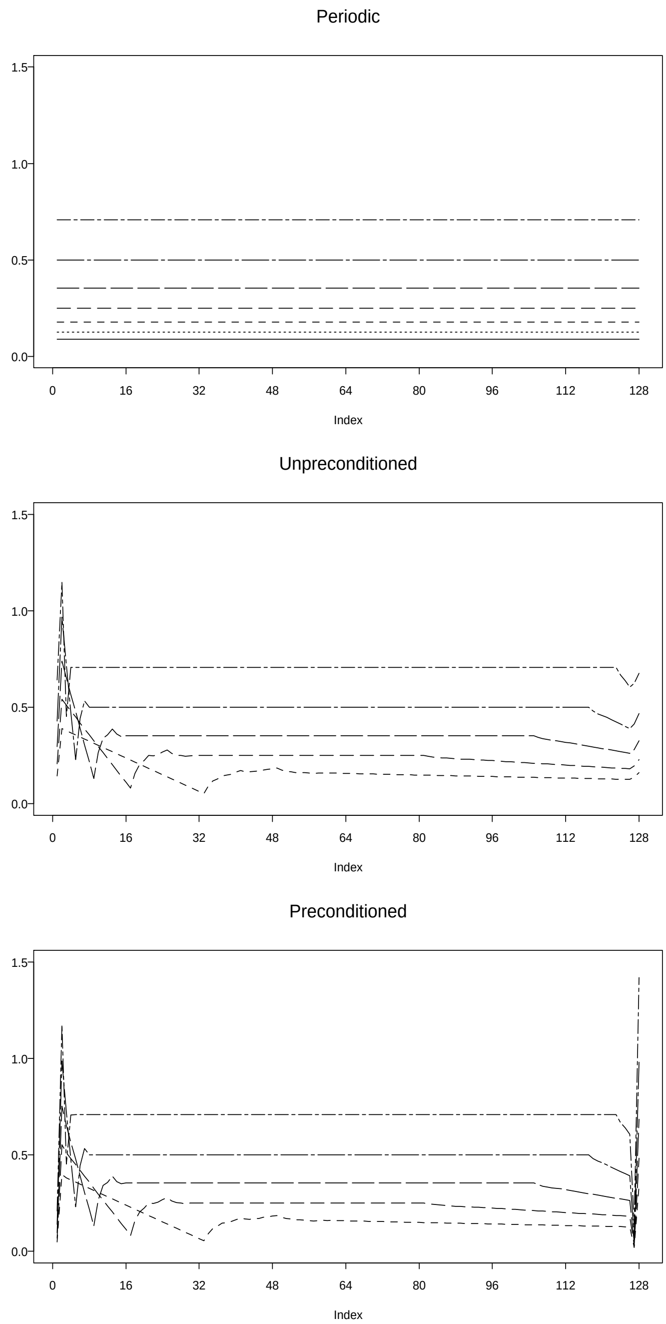


Figure 6.10: Weights attached to original data to give $\sum_{k=1}^{N/2m} x_k^m$ (sum of the smooths at scale m)

fractionally differenced process. This contrasts to the method of Kaplan and Kuo (1993) and Tewfik and Kim (1992) who estimate the correlations via simulation (with the exception of the Haar wavelet).

Figures 6.11 to 6.13 show the correlation structure for the discrete wavelet transform coefficients employing a variety of wavelets, with $N = 32$ and $d = 0.4$. These figures demonstrate that there are no correlations within any of the scales m. There exist some small correlations between scales, shown by the lower peaks. The largest correlations occur at the ends of contiguous scales, an effect caused by the periodic boundary conditions and can be seen in Figures 6.12 and 6.13. This can be appreciated by comparing Figure 6.11, which shows the correlations for the Haar wavelet, to the alternative wavelets, and noticing that none of the peaks present in the other cases are present here. This is due to the fact that there are no end-effects for Haar, the support always being internal to the finite data length.

Also notice that the between scale correlations reduce in size as the number of filter coefficients in the wavelet increase, but of course this is countered by the fact that the boundary effects become more severe.

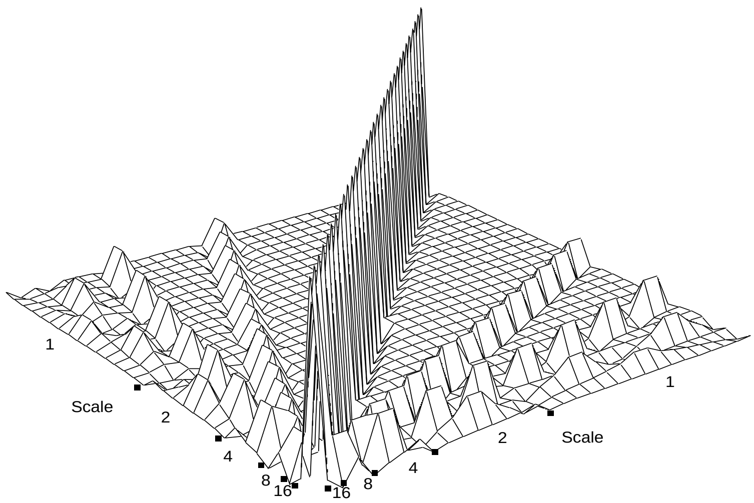


Figure 6.11: Correlations of DWT coefficients: Haar

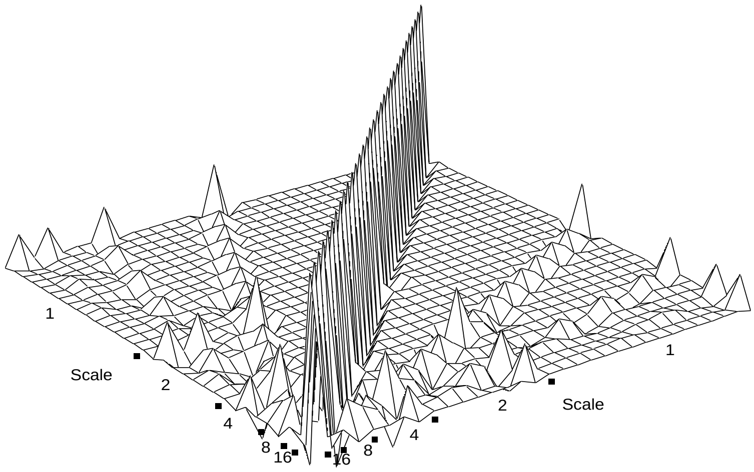


Figure 6.12: Correlations of DWT coefficients: EP(4)

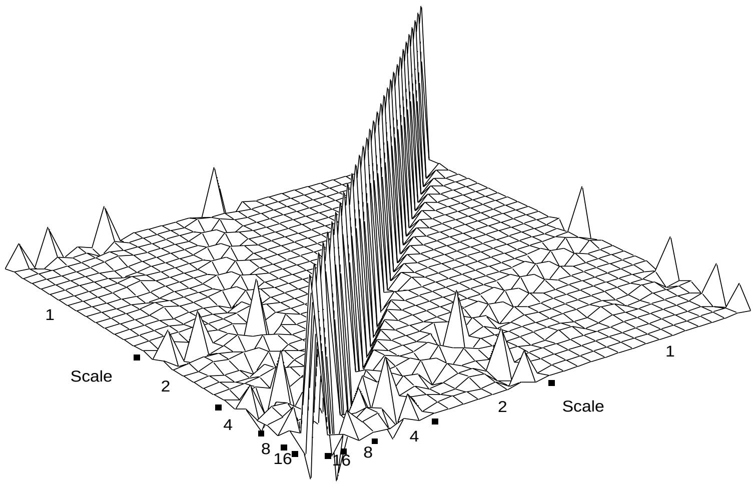


Figure 6.13: Correlations of DWT coefficients: LA(8)

6.6.2 Correlations of Haar transform coefficients

The method of calculating the correlations outlined in the previous section is computationally very intensive and can involve large matrix operations, as well as the calculation of the wavelet matrix W.

A more efficient iterative method is available, which employs a method similar in flavour to that used by Kaplan and Kuo (1993), which will now be described.

For the Haar wavelet, the following relationships hold between the detail and smooth coefficients at progressive scales

$$x_j^{2m} = \frac{1}{\sqrt{2}}(x_{2j-1}^m + x_{2j}^m) \quad m \geq 1.$$

and

$$d_j^{2m} = \frac{1}{\sqrt{2}}(x_{2j-1}^m - x_{2j}^m) \quad m \geq 1.$$

where x_j^m represents the 'smooth' coefficients, and d_j^m represent the 'detail' coefficients at scale $m = 1, 2, 4, \ldots, N/2$ and within the mth scale $j = 1, 2, 3, \ldots N/2m$.

Letting s_τ^m represent the autocovariance of $\{x_j^m\}$, and $\{x_j\}$ be the original data (with autocovariance s_τ^0), with the first step in the transform defined as

$$x_j^1 = \frac{1}{\sqrt{2}}(x_{2j-1} + x_{2j}),$$

and

$$d_j^1 = \frac{1}{\sqrt{2}}(x_{2j-1} - x_{2j}).$$

Then

$$\begin{aligned}
s_\tau^{2m} &= E(x_t^{2m} x_{t+\tau}^{2m}) \\
&= E\left\{\tfrac{1}{\sqrt{2}}(x_{2t-1}^m + x_{2t}^m)\tfrac{1}{\sqrt{2}}(x_{2t+2\tau-1}^m + x_{2t+2\tau}^m)\right\} \\
&= \tfrac{1}{2}\left\{E(x_{2t-1}^m x_{2t+2\tau-1}^m) + E(x_{2t-1}^m x_{2t+2\tau}^m) + E(x_{2t}^m x_{2t+2\tau-1}^m) + E(x_{2t}^m x_{2t+2\tau}^m)\right\} \\
&= \tfrac{1}{2}\left\{s_{2\tau}^m + s_{2\tau+1}^m + s_{2\tau-1}^m + s_{2\tau}^m\right\}.
\end{aligned}$$

Similarly, defining $s_{\tau,d}^m$ as the autocovariance of $\{d_j^m\}$, then

$$\begin{aligned}
s_{\tau,d}^{2m} &= E(d_t^{2m} d_{t+\tau}^{2m}) \\
&= E\left\{\tfrac{1}{\sqrt{2}}(x_{2t}^m - x_{2t-1}^m)\tfrac{1}{\sqrt{2}}(x_{2t+2\tau}^m - x_{2t+2\tau-1}^m)\right\} \\
&= \tfrac{1}{2}\left\{E(x_{2t-1}^m x_{2t+2\tau-1}^m) - E(x_{2t-1}^m x_{2t+2\tau}^m) - E(x_{2t}^m x_{2t+2\tau-1}^m) + E(x_{2t}^m x_{2t+2\tau}^m)\right\} \\
&= \tfrac{1}{2}\left\{s_{2\tau}^m - s_{2\tau+1}^m - s_{2\tau-1}^m + s_{2\tau}^m\right\}.
\end{aligned}$$

So we have

$$\begin{aligned} s_\tau^{2m} &= s_{2\tau}^m + \tfrac{1}{2}\left\{s_{2\tau+1}^m + s_{2\tau-1}^m\right\} \\ s_{\tau,d}^{2m} &= s_{2\tau}^m - \tfrac{1}{2}\left\{s_{2\tau+1}^m + s_{2\tau-1}^m\right\} \qquad \text{for} \quad m \geq 1. \end{aligned}$$

and

$$\begin{aligned} s_\tau^1 &= s_{2\tau} + \tfrac{1}{2}\left\{s_{2\tau+1} + s_{2\tau-1}\right\} \\ s_{\tau,d}^1 &= s_{2\tau} - \tfrac{1}{2}\left\{s_{2\tau+1} + s_{2\tau-1}\right\}. \end{aligned}$$

This provides us with an algorithm for calculating the autocovariance sequence of the discrete wavelet transform coefficients, when using the Haar wavelet. This algorithm is described for the case of a fractionally differenced white noise process in Table 6.2. We use the fact that, for a fractionally differenced process, the following relationship holds for successive elements in the autocovariance sequence of the stationary process $\{X_t\}$:

$$s_\tau = s_{\tau-1}\left(\frac{\tau+d-1}{\tau-d}\right).$$

6.7 Phase of wavelet filters

As we intend to use the wavelet transform as an analysis tool, we must be aware of its properties. An important measure for any filter is its phase function, with this we can establish whether there are any inherent properties of the filter such as phase shifts or rotations which could effect our interpretation of the results.

In this section we shall establish the phase of the filters which are applied to the original data to obtain the discrete wavelet transform coefficients and the smooth coefficients at each of the scales. These filters, shall be denoted $\{g_j^m\}$ and $\{h_j^m\}$ for $j = 0, \ldots, L_m - 1$ at scale m, where $L_m = (2m-1)L_1 - 2m + 2$, and L_1 is the length of the filters applied at the first scale. Notice that, for all the wavelets of interest here, L_1 and consequently L_m are always even. These low and high-pass filters are iterated versions of the original high and low-pass filters $\{g_j\}$ and $\{h_j\}$ for $j = 0, \ldots, L_1 - 1$, and the iteration is given by

$$g_j^m = \sum_{k \in \mathbf{Z}} g_k^{m/2} g_{j-2k},$$

1. set $s_0^0 = \frac{(-2d)!}{[(-d)!]^2}$

2. for $\tau = 1, \ldots, N-1$

$$s_\tau^0 = s_{\tau-1}^0 \frac{\tau+d-1}{\tau-d}$$

endfor

3. for $j = 1, 2, \ldots, \log_2(N)$

set $m = 2^{j-1}$

for $\tau = 0, \ldots, \frac{N}{2m} - 1$

$$s_\tau^j = s_{2\tau}^{j-1} + \frac{1}{2}\left(s_{2\tau+1}^{j-1} + s_{2\tau-1}^{j-1}\right)$$

$$s_{\tau,d}^j = s_{2\tau}^{j-1} - \frac{1}{2}\left(s_{2\tau+1}^{j-1} + s_{2\tau-1}^{j-1}\right)$$

endfor

endfor

Table 6.2: Algorithm for calculating acvs of Haar transform coefficients

$$h_j^m = \sum_{k \in \mathbf{Z}} h_k^{m/2} g_{j-2k}.$$

The filters associated with Daubechies' least asymmetric wavelets are designed to have approximately zero phase, if they are applied with a shift of half the filter length. Using relationships between the transfer functions of the filters, we will establish the shifts required for the filters applied at scale m to maintain zero phase, and show that, if the original filter length (L_1) is even when divided by 2, these shifts are again half the length of the filter applied at that scale ($L_m/2 = (2m-1)L_1/2 - m + 1$).

First, we shall define the phase of a filter given its transfer function. Let us define a digital filter L to be such that

$$L\{\{X_t\}\} = \{Y_t\}.$$

It then follows (e.g. Percival and Walden (1993, p156)) that

$$L\{\{e^{i2\pi ft}\}\} = G(f)\{e^{i2\pi ft}\},$$

where $G(\cdot)$ defines the transfer function.

In general $G(\cdot)$ is a complex-valued function, so we can write

$$G(f) = |G(f)|e^{i\theta(f)},$$

where $|G(f)|$ and $\theta(f)$ are called, respectively, the gain function and the phase function.

Before we proceed to calculate the phase of the wavelet filters, we must establish some basic filter relationships.

6.7.1 Transfer function of combined filters

As the discrete wavelet transform can be thought of as a filter bank involving the iteration of high and low-pass filters, to establish the phase of these iterated filters, we must first establish the relationship between the transfer functions of combined filters.

Given two filters $\{h_t\}$ of length L, and $\{g_t\}$ of length K. If we define

$$y_t = h * x_t \quad \text{and} \quad z_t = g * y_t,$$

where $*$ denotes convolution, then we can write

$$z_t = c * x_t,$$

where $\{c_t\}$ is obtained from the convolution of $\{h_t\}$ and $\{g_t\}$, thus

$$c_t = h * g_t.$$

If we let $H(\cdot)$ and $G(\cdot)$ represent the transfer functions for $\{h_t\}$ and $\{g_t\}$,

$$H(f) = \sum_{t=0}^{L-1} h_t e^{-i2\pi f t} \quad \text{and} \quad G(f) = \sum_{t=0}^{K-1} g_t e^{-i2\pi f t}. \tag{6.13}$$

Then the transfer function $C(\cdot)$ for $\{c_t\}$ is given by

$$C(f) = H(f)G(f). \tag{6.14}$$

We shall use this relationship to establish the phase of the combined filters.

6.7.2 Phase and transfer functions of advanced filters

As the least asymmetric wavelets have filters designed to have zero phase after a time shift, we must establish the relationship between the phase of our original filters and the phase of the filter after a time shift.

Using the same definitions as in the previous section that $y_t \equiv \sum_{k=0}^{L-1} h_k x_{t-k}$ and that $H(\cdot)$ is the transfer function for $\{h_k\}$. If we advance the output by τ units giving

$$\tilde{y}_{t-\tau} \equiv \sum_{k=0}^{L-1} h_k x_{t-k}$$

or equivalently

$$\tilde{y}_t = \sum_{k=0}^{L-1} h_k x_{t+\tau-k},$$

then the transfer function for the advance filter is

$$\tilde{H}(f) = e^{i2\pi f\tau} H(f). \tag{6.15}$$

Writing $H(f) = e^{i\theta(f)}|H(f)|$, then we can also write $\tilde{H}(f) = e^{i(\theta(f)+2\pi f\tau)}|H(f)|$, so that the phase of the advanced filter is

$$\tilde{\theta}(f) \equiv \theta(f) + 2\pi f\tau. \tag{6.16}$$

6.7.3 Transfer function of decimated filters

At each stage in the discrete wavelet transform, we filter the data, and only take every other point in the output. At a particular scale m, we can consider the output as an application of the filters on the original data, but only taking every $2m$th point of the output. We must therefore establish the relationship between the transfer function of the original filter, and the transfer function of a filter whose output only consists of every mth point.

Let us form a new filter $\{\tilde{h}_j\}$ for $j = 0, \ldots, m(L-1)$, formed by inserting $m-1$ zeroes in the between each element of the current filter $\{h_j\}$ as follows

$$
\begin{array}{cccc}
\tilde{h}_0 = h_0 & \tilde{h}_1 = 0 & \ldots & \tilde{h}_{m-1} = 0 \\
\tilde{h}_m = h_1 & \tilde{h}_{m+1} = 0 & \ldots & \tilde{h}_{2m-1} = 0 \\
\tilde{h}_{2m} = h_2 & \tilde{h}_{2m+1} = 0 & \ldots & \tilde{h}_{3m-1} = 0 \\
\vdots & \vdots & & \vdots \\
\tilde{h}_{m(L-2)} = h_{L-2} & \tilde{h}_{m(L-2)+1} = 0 & \ldots & \tilde{h}_{m(L-1)-1} = 0 \\
\tilde{h}_{m(L-1)} = h_{L-1}. & & &
\end{array}
$$

That is

$$
\tilde{h}_j = \begin{cases} h_{j/m} & \text{if } j \text{ divides } m \\ 0 & \text{otherwise} \end{cases}
$$

We can calculate the transfer function for $\{\tilde{h}_j\}$ as

$$
\tilde{G}(f) \equiv \sum_{j=0}^{m(L-1)} \tilde{h}_j e^{-i2\pi fj} = \sum_{j=0}^{L-1} h_j e^{-i2m\pi fj} = G(mf). \tag{6.17}
$$

6.8 Phase of wavelet filters at each scale

Now we have established all the filter relationships required to establish the phase of the filters applied at each scale.

Let $\{h_k^1\}$ for $k = 0, \ldots, L_1 - 1$ represent the high-pass filter at the first scale, and $\{g_k^1\}$ for $k = 0, \ldots, L_1 - 1$ be the corresponding low-pass filter (previously referred to as $\{h_k\}$ and $\{g_k\}$). Recall that the quadrature mirror relationship holds:

$$
g_k^1 = (-1)^{k+1} h_{L_1-k-1}^1 \quad \text{for} \quad k = 0, \ldots, L_1 - 1.
$$

We assume that the filters are applied without an advance.

Letting $H_1(\cdot)$ and $G_1(\cdot)$ be the transfer functions corresponding to the filters $\{h_k^1\}$ and $\{g_k^1\}$. We can write

$$H_1(f) = |H_1(f)|e^{i\theta_1^{(H)}(f)} \quad \text{and} \quad G_1(f) = |G_1(f)|e^{i\theta_1^{(G)}(f)}.$$

Where $\theta_1^{(H)}(\cdot)$ and $\theta_1^{(G)}(\cdot)$ are the phase functions for the filters $\{h_k^1\}$ and $\{g_k^1\}$. We have

$$\begin{aligned}
G_1(f) &= \sum_{k=0}^{L_1-1} g_k^1 e^{-i2\pi fk} = \sum_{k=0}^{L_1-1} (-1)^{k+1} h_{L_1-k-1} e^{-i2\pi fk} \\
&= \sum_{k=0}^{L_1-1} e^{i\pi(k+1)} h_{L_1-k-1} e^{-i2\pi fk} \\
&= e^{i\pi} \sum_{k=0}^{L_1-1} h_{L_1-k-1} e^{k(-i2\pi f+i\pi)} \\
&= e^{i\pi} \sum_{k=0}^{L_1-1} h_k e^{(L_1-k-1)(-i2\pi f+i\pi)} \\
&= e^{i\pi} e^{-(L_1-1)(2\pi fi-i\pi)} \sum_{k=0}^{L_1-1} h_k e^{-k(-i2\pi f+i\pi)} \\
&= e^{i\pi L_1} e^{-i2\pi f(L_1-1)} \sum_{k=0}^{L_1-1} h_k e^{-i2\pi k(\frac{1}{2}-f)} \\
&= e^{-i2\pi f(L_1-1)} H_1\left(\frac{1}{2}-f\right),
\end{aligned}$$

as L_1 is even, from which we derive

$$e^{i\theta_1^{(G)}(f)}|G_1(f)| = e^{-i2\pi f(L_1-1)} e^{i\theta_1^{(H)}(\frac{1}{2}-f)} \left|H_1\left(\frac{1}{2}-f\right)\right|,$$

giving us the relationship between the phases of the high and low-pass filters at the first scale:

$$\theta_1^{(G)}(f) = \theta_1^{(H)}\left(\frac{1}{2}-f\right) - 2\pi f(L_1-1). \tag{6.18}$$

Daubechies' class of least asymmetric (LA) filters is defined so that the wavelet filter has approximately zero phase over all frequencies, provided we advance its output by $\tau = L_1/2$ units, that is letting

$$w_t^1 = \sum_{k=0}^{L_1-1} h_k^1 x_{t-k+(L_1/2)}$$

define the output of the wavelet filter. Applying Equation 6.16, this means

$$\theta_1^{(H)}(f) \approx -\pi f L_1,$$

where $\theta_1^{(H)}(f)$ refers to the phase of the unadvanced wavelet filter. Substituting this into Equation 6.18 gives

$$\begin{aligned} \theta_1^{(G)}(f) &\approx -\pi\left(\tfrac{1}{2}-f\right)L_1 - 2\pi f(L_1-1) \\ &= -\pi\tfrac{L_1}{2} - 2\pi f\left(\tfrac{L_1}{2}-1\right). \end{aligned}$$

If $L_1/2$ is even, we have, modulo 2π,

$$\theta_1^{(G)}(f) \approx -2\pi f\left(\frac{L_1}{2}-1\right).$$

Notice that, if $L_1/2$ is odd the filter will have a phase shift, to avoid this, instead of advancing the filter by $L_1/2$ we advance it by $\tau = L_1/2 \pm 1$, which, on using Equation 6.16 gives

$$\theta_1^{(H)}(f) \approx -2\pi f\left(\frac{L_1}{2} \pm 1\right).$$

Substituting this into Equation 6.18 gives

$$\begin{aligned} \theta_1^{(G)}(f) &\approx -2\pi\left(\tfrac{1}{2}-f\right)\left(\tfrac{L_1}{2}\pm 1\right) - 2\pi f(L_1-1) \\ &\approx -\pi\left(\tfrac{L_1}{2}\pm 1\right) - 2\pi f\left(\tfrac{L_1}{2}\mp 1 - 1\right). \end{aligned}$$

As $L_1/2 \pm 1$ is even, we have, modulo 2π,

$$\theta_1^{(G)}(f) \approx -2\pi f\left(\frac{L_1}{2} \mp 1 - 1\right).$$

This gives us the phase functions for the first scale. Recall from the definition of the discrete wavelet transform that the detail coefficients at scale 2 are the result of decimating by two the output from the first low-pass filter, and then high-pass filtering. These actions can be combined to provide appropriate filters that would act on the original series to produce the same results. Using Equations 6.14 and 6.17, the transfer functions of the filters at scale two are given by

$$\begin{aligned} H_2(f) &= G_1(f)H_1(2f), \\ G_2(f) &= G_1(f)G_1(2f). \end{aligned}$$

The phase function for the bandpass filter at scale two is thus

$$\theta_2^{(H)}(f) = \theta_1^{(G)}(f) + \theta_1^{(H)}(2f).$$

First, we take the case of $L_1/2$ even, and using the fact that $L_2 = 3L_1 - 2$, then

$$\begin{aligned}
\theta_2^{(H)}(f) &\approx -2\pi f\left(\tfrac{L_1}{2} - 1\right) - 2\pi f L_1 \\
&= -2\pi f\left(\tfrac{3L_1}{2} - 1\right) \\
&= -2\pi f\left(\tfrac{L_2}{2}\right).
\end{aligned}$$

In the case of $L_1/2$ odd, then

$$\begin{aligned}
\theta_2^{(H)}(f) &\approx -2\pi f\left(\tfrac{L_1}{2} \mp 1 - 1\right) - 4\pi f\left(\tfrac{L_1}{2} \pm 1\right) \\
&= -2\pi f\left(\tfrac{3L_1}{2} \pm 1 - 1\right) \\
&= -2\pi f\left(\tfrac{L_2}{2} \pm 1\right).
\end{aligned}$$

Hence, if we were to advance the filter by half its length if the filter length is even when divided by two, or half its length plus or minus one otherwise, we would obtain an approximately zero phase filter. In fact, we shall now show that, whatever the original advance is to obtain an approximately zero phase filter, this advance propagates in a simple fashion to all higher scale filters.

Let us assume that a shift of $L_1/2 + c$, which is even, produces an approximately zero phase filter for scale one, that is from Equation 6.16

$$\theta_1^{(H)}(f) = -2\pi f\left(\frac{L_1}{2} + c\right).$$

We now show that

$$\theta_m^{(H)}(f) = -2\pi f\left(\frac{L_m}{2} + c\right).$$

Implying that the filter will be zero phase if shifted by $L_m/2 + c$.

Proof

Recall that

$$\begin{aligned}
\theta_1^{(G)}(f) &= \theta_1^{(H)}(0.5 - f) - 2\pi f(L_1 - 1) \\
&= -\pi\left(\tfrac{L_1}{2} + c\right) + 2\pi f\left(\tfrac{L_1}{2} + c\right) - 2\pi f(L_1 - 1) \\
&= -\pi\left(\tfrac{L_1}{2} + c\right) - 2\pi f\left(\tfrac{L_1}{2} - (c+1)\right) \\
&= -2\pi f\left(\tfrac{L_1}{2} - (c+1)\right),
\end{aligned}$$

modulo 2π, and using our assumption that $(L_1/2 + c)$ is even. Let $m = 2^M$, then

$$H_m(f) = G_{m/2}(f)H_1(mf) = H_1(mf)\prod_{j=0}^{M-1} G_1(2^j f),$$

and

$$G_m(f) = G_{m/2}(f)G_1(mf) = \prod_{j=0}^{M} G_1(2^j f).$$

Giving the following phase function for the wavelet filter at scale m

$$\begin{aligned}
\theta_m^{(H)} &= \theta_1^{(H)}(2^M f) + \sum_{j=0}^{M-1} \theta_1^{(G)}(2^j f) \\
&= -2\pi f \left\{ \frac{L_1}{2} + c \right\} m - 2\pi f \left\{ \frac{L_1}{2} - (c+1) \right\} \sum_{j=0}^{K-1} 2^j \\
&= -2\pi f \left\{ \frac{L_1}{2} m + mc + \left(\frac{L_1}{2} - (c+1) \right) (m-1) \right\} \\
&= -2\pi f \left\{ L_1 \left(\frac{2m-1}{2} \right) - m + 1 + c \right\} \\
&= -2\pi f \left(\frac{L_m}{2} + c \right),
\end{aligned}$$

and the result follows.

6.9 FFT calculation of Phase

From Equation 6.13 it immediately follows that the transfer function $G(f)$ is the Fourier transform of the impulse response sequence.

We can use this definition to exploit the use of the Fast Fourier Transform (FFT), which calculates the Fourier transform at the Fourier frequencies: $0, 1/N, \ldots, 1/2$, where N is the length of the series being transformed.

So, in order to calculate the phase at as finer grid of frequency points as is desired, we can pad the filter with zeros and use the algorithm outlined in Table 6.3. Here `atan2` is the two variable arctan function (available in C or Fortran), whose output range is $[-\pi, \pi]$, and S_m is the required shift at each scale, as defined in Table 6.4.

6.10 Phase of least asymmetric wavelets

We shall now examine how close to zero phase the least asymmetric wavelets are, using the shifts outlined in Table 6.4. The LA(k) wavelets for $k = 8, 12, 16$ and 20 are shown in Figure 6.14, while the remainder are shown in Figure 6.15. We can

1. Let $\{h_{t,m}\}$ define the filter at scale m:

$$h_{t,m} = \begin{cases} h_{t,m} & \text{if } 0 < t < L_m - 1 \\ 0 & \text{if } L < t < N - 1. \end{cases}$$

2. Calculate the Fourier transform of $\{h_{t,m}\}$ via the *FFT*

$$Y(f_i) = \sum_{t=0}^{N-1} h_{t,m} e^{-i2\pi f_i t}.$$

where $f_i = i/N$.

3. The transfer function is then

$$G(f_i) = e^{i2\pi f_i S_m} Y(f_i),$$

where S_m is defined in Table 6.4

4. From the transfer function we can obtain the phase of the filter at scale m:

$$\theta_m(f_i) = \mathtt{atan2}\left(\Im(G(f_i)), \Re(G(f_i))\right).$$

Table 6.3: Algorithm: *FFT* calculation of phase

Wavelet	S_m
LA(8)	$L_m/2$
LA(10)	$L_m/2-1$
LA(12)	$L_m/2$
LA(14)	$L_m/2+1$
LA(16)	$L_m/2$
LA(18)	$L_m/2-1$
LA(20)	$L_m/2$

Table 6.4: Shifts required for zero phase at progressive scales

see that the LA wavelets do maintain approximately zero phase across all scales, though those that have lengths that are even when divided by two appear to approximate zero phase filters with more accuracy than the others.

For comparison Figure 6.16 shows the phase of the Haar wavelet, which is approximately linear over passbands. While the slope can be removed by a time shift, if a time series were analyzed using the Haar wavelet, it may be difficult to line up events with the coefficients due to the phase properties of the filter. While the Haar wavelet is appealing in its simplicity, we must be aware of problems that may occur with its use, both in terms of poor frequency resolution and phase shifts.

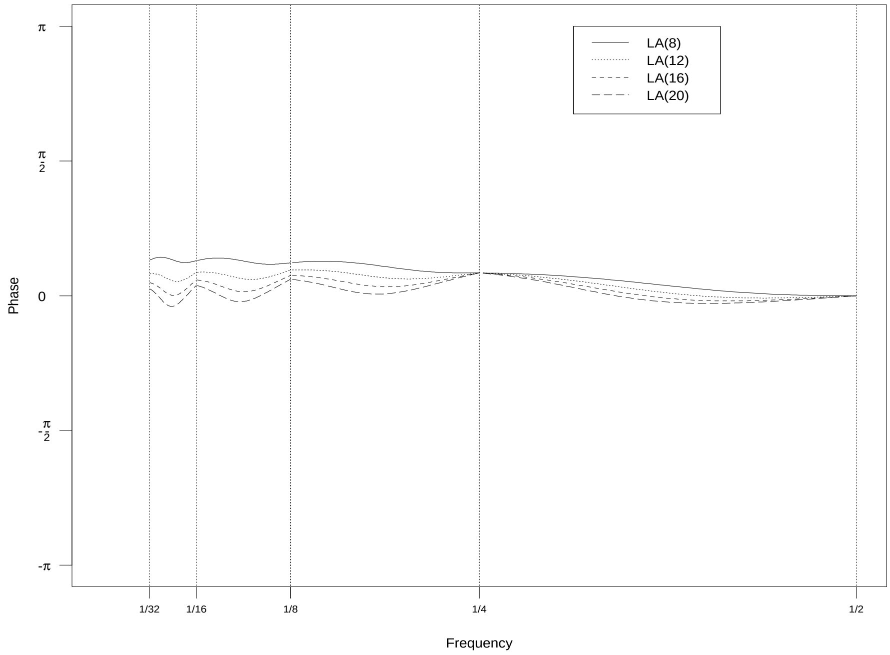


Figure 6.14: Phases of LA wavelets

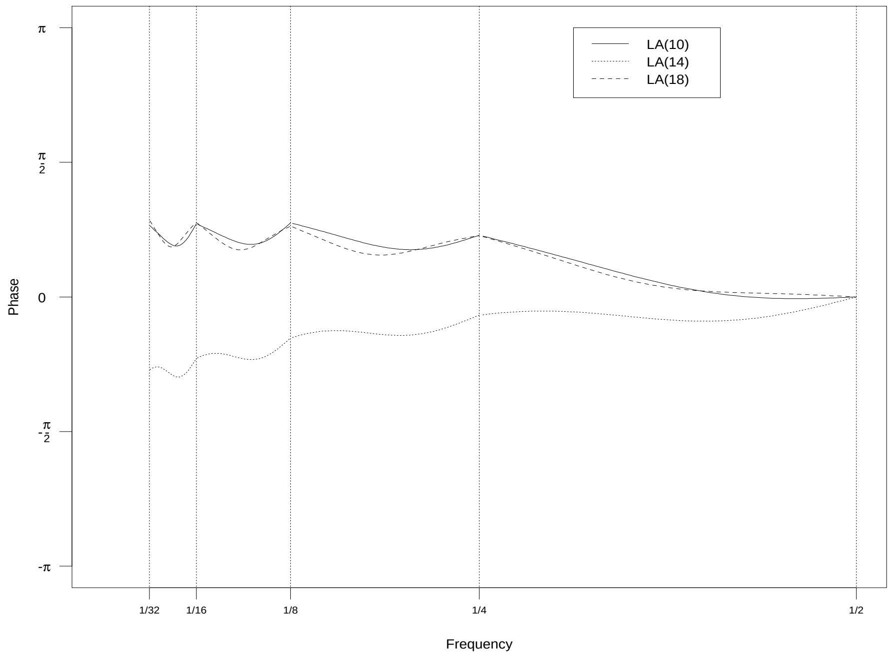


Figure 6.15: Phases of LA wavelets

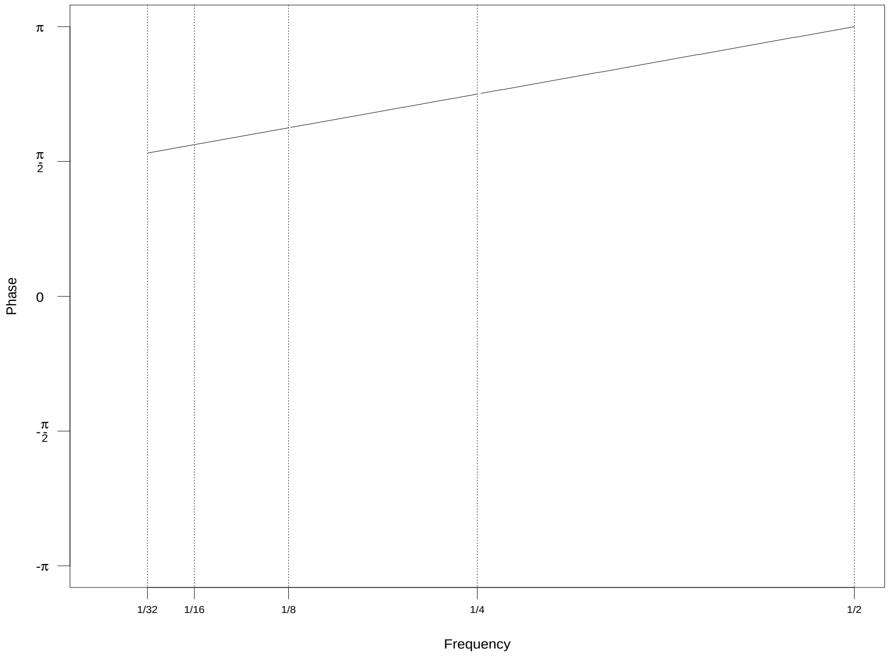


Figure 6.16: Phase of Haar wavelet

Chapter 7

The Analysis and Synthesis of Long Memory Processes using Wavelets

7.1 Introduction

The usefulness of wavelets as a tool for analyzing long memory processes has recently been investigated in a series of publications by Wornell (1990,91,93) and Wornell & Oppenheim (1992a,b), Abry *et al* (1993), Tewfik *et al* (1993) and Percival & Guttorp (1994).

In this chapter we shall develop a new method of analyzing and synthesising discrete time long memory processes via wavelets.

7.2 Bandpass variance

Firstly considering a stationary process which is exactly of $1/f$-type over a number of octaves of frequency. The bandpass variance (Rutman 1978) B_m for a spectrum $S(f) = \sigma_\epsilon^2 f^{-\gamma}$ in the frequency octave $(1/4m, 1/2m]$ is

$$B_m = \int_{1/4m}^{1/2m} \sigma_\epsilon^2 f^{-\gamma}\, df = \frac{\sigma_\epsilon^2 (2m)^{\gamma-1}[1-2^{\gamma-1}]}{1-\gamma}.$$

Instead of the true spectrum over a bandpass, we could approximate the spectrum with a constant power spectrum of value S_m. Then the bandpass variance would be:

$$\sigma_m^2 = \int_{1/4m}^{1/2m} S_m \, df = \frac{S_m}{4m}.$$

In order to maintain equal bandpass variances let

$$S_m = \sigma_\epsilon^2 (2m)^\gamma [2 - 2^\gamma]/[1 - \gamma], \tag{7.1}$$

so that $S_m \propto (2m)^\gamma$. Then the ratios of the bandpass variances from one scale to the next are independent of scale:

$$\frac{S_m}{S_{2m}} = 2^{-\gamma}.$$

Figure 7.1 shows $\log_2 S_m$ plotted against $3/8m$ (the centre frequency of the passband for scale m, on a $\log_2$ scale), when $\gamma = 0.8$. This produces a straight line with gradient $-\gamma = 0.8$. By definition of the wavelet transform, the de-

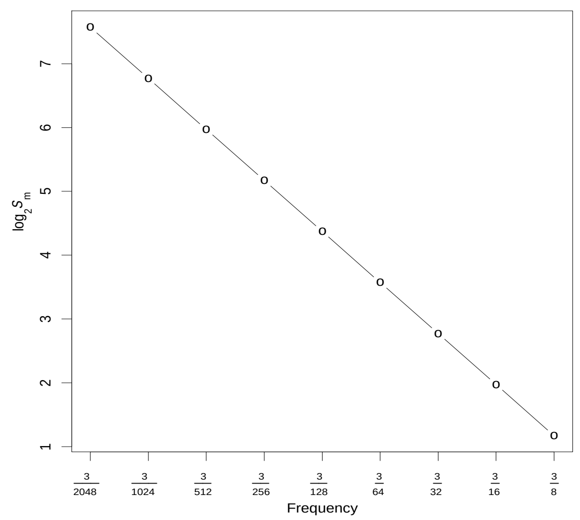


Figure 7.1: $\log_2 S_m$ against centre frequency

tail coefficients in the octave band $(1/4m, 1/2m]$ are a passband representation of the original process; the variance of the detail coefficients for the passband $(1/4m, 1/2m]$ is just S_m, for $m = 1, 2, 4, \ldots, N/2$.

We can exploit this relationship between the variance of the detail coefficients of the discrete wavelet transform of a long memory process in the analysis and synthesis of such processes.

The discrete wavelet transform, as well as containing the detail coefficients, also contains one 'smooth' coefficient – the result of applying many low-pass filters to the original data. In order to synthesize any process we must calculate the variance of this coefficient. The bandpass variance in the frequency band $(0, 1/2N]$ is

$$B_N = \int_0^{1/2N} \sigma_\epsilon^2 f^{-\gamma}\, df = c(2N)^{\gamma-1}/(1-\gamma).$$

Hence

$$S_N = \sigma_\epsilon^2 (2N)^\gamma/(1-\gamma). \tag{7.2}$$

7.3 Synthesis of 1 -type processes

We can derive the expected value and variance of the output from the inverse wavelet transform in a similar way to the forward discrete wavelet transform as described in Section 6.6. The inverse wavelet transform can also be written as a matrix equation

$$X_N = W^{-1}Z = W^T Z,$$

assuming that W is orthogonal, as is the case for all wavelets with periodic boundary conditions employed here, giving

$$E\{X_N\} = W^T E\{Z\} \text{ and } \mathrm{var}\{X_N\} = W^T \mathrm{var}\{Z\} W. \tag{7.3}$$

We have already seen that the discrete wavelet transform of a $1/f$-type process has detail coefficients that are nearly uncorrelated both between and within scales, if we ignore these small correlations, and assume that $\mathrm{var}\{Z\}$ can be written as follows

$$\text{var}\{Z\} = \begin{pmatrix} S_N & & & & \\ & S_{N/2} & & & \\ & & \begin{pmatrix} S_{N/4} & \\ & S_{N/4} \end{pmatrix} & & \\ & & & \ddots & \\ & & & & \begin{pmatrix} S_1 & & \\ & \ddots & \\ & & S_1 \end{pmatrix} \end{pmatrix}, \quad (7.4)$$

then we can synthesize such a process by way of the algorithm described in Table 7.1.

A similar method of synthesizing $1/f$-type processes has been derived by Wornell (1990), with fuller details in Wornell (1991) and (1993). The algorithm he describes differs from the method described here in that he gives the variance of the detail coefficients up to proportionality, and does not offer advice on how to simulate the one 'smooth' coefficient (part 2 of our algorithm). Wornell shows that his algorithm produces a series with a nearly $1/f$ spectrum, which he defines as having spectrum of the form

$$S(f) = \begin{cases} \frac{1}{f^\gamma} & \text{for } f_1 < f < f_2 \\ 0 & \text{otherwise,} \end{cases}$$

that is, a spectrum that is $1/f$-type over a defined frequency passband. Here, we are interested in discrete-time stationary processes; the inclusion of the one 'smooth' coefficient is vital for the success of the method.

1. Generate d_j^m from $\mathrm{N}(0, S_m)$ for $m = 1, 2, 4, \ldots, N/2$, and $j = 1, 2, \ldots, N/2m$, with S_m as defined in Equation 7.1.
2. Generate $x_1^{N/2}$ from $\mathrm{N}(0, S_N)$, with S_N as defined in Equation 7.2.
3. Form synthesized vector $$Z = \left(x_1^{N/2}, d_1^{N/2}, d_1^{N/4}, d_2^{N/4}, d_1^{N/8}, \ldots, d_4^{N/8}, \ldots, d_1^1, \ldots, d_{N/2}^1\right)^T.$$
4. Form the inverse wavelet transform of Z to obtain the synthesized $1/f$-type process $$X = \{x_1, x_2, \ldots, x_N\}.$$

Table 7.1: Algorithm: Synthesis of $1/f$-type processes via the discrete wavelet transform

7.4 Synthesis of general long memory processes

In the previous section we assumed that the process we wished to synthesize was of perfect $1/f$-type, any other similar process can be generated in the same way once the variance progression in the octave bands have been calculated (i.e. those defined in Equation 7.1 and 7.2), and a check has been made to ensure that the correlation matrix of the transformed series has small off-diagonal terms (i.e. as shown in Figures 6.11 to 6.13). In the following section we shall demonstrate how to synthesize a fractionally differenced process using just such a method.

7.4.1 Fractionally differenced process

An exact method of simulating fractionally differenced processes has been described by Hosking (1984), so we can compare the results of our simulations to his method. However, in the case where one wishes to generate a particular 'near to $1/f$ process' an exact method will generally not be available.

Suppose we wish to simulate a series of length $N = 2^p = 128$ ($p = 7$), we first need to calculate the bandpass variances $B_m, m = 1, 2, 4, \ldots, 64$. For a fractionally differenced process these are defined as

$$B_m = \sigma_\epsilon^2 \int_{1/4m}^{1/2m} (2\sin(\pi f))^{-2d}\, df \quad \text{for} \quad m = 1, 2, 4, \ldots, 64.$$

These bandpass variances were calculated numerically for several values of d. To establish the bandpass variance in the frequency interval $(0, 1/256]$ (i.e. B_N) we subtract the sum of the bandpass variances from half the true variance of the process (which is, by definition the integral of the spectrum in the frequency range $(0, 1/2]$, namely $\sigma_\epsilon^2\Gamma(1-2d)/2\Gamma^2(1-d)$). To give the results shown in Table 7.2. Notice that as d increases the bandpass variance in the lowest frequency

m	bandpass	B_m/σ_ϵ^2			
		d=0.05	d=0.25	d=0.4	d=0.45
1	$(\frac{1}{4}, \frac{1}{2}]$	0.235851	0.235851	0.157326	0.148548
2	$(\frac{1}{8}, \frac{1}{4}]$	0.123975	0.123975	0.118052	0.117377
4	$(\frac{1}{16}, \frac{1}{8}]$	0.066135	0.066135	0.099274	0.105364
8	$(\frac{1}{32}, \frac{1}{16}]$	0.035401	0.035401	0.085690	0.097377
16	$(\frac{1}{64}, \frac{1}{32}]$	0.001897	0.018966	0.074439	0.090641
32	$(\frac{1}{128}, \frac{1}{64}]$	0.001016	0.010163	0.064769	0.084521
64	$(\frac{1}{256}, \frac{1}{128}]$	0.000545	0.005446	0.049078	0.078850
128	$(0, \frac{1}{256}]$	0.006288	0.049868	0.379122	1.098537

Table 7.2: Bandpass variances for an fdp(d)

band increases, demonstrating the preponderance of energy at low frequencies for fractionally differenced processes with a high value of d. Having established the

B_m's we can calculate the S_m's $(= 4mB_m)$ and then substitute these into the algorithm described in Table 7.1.

Figure 7.2 shows $\log_2 S_m$ plotted against $3/8m$ (the centre frequency of the passband for scale m, on a $\log_2$ scale), for fractionally differenced processes with varying values of d, and $\sigma_\epsilon^2 = 1$, we can see that these values increase as the frequency decreases in a similar way to $1/f$ processes (see Figure 7.1).

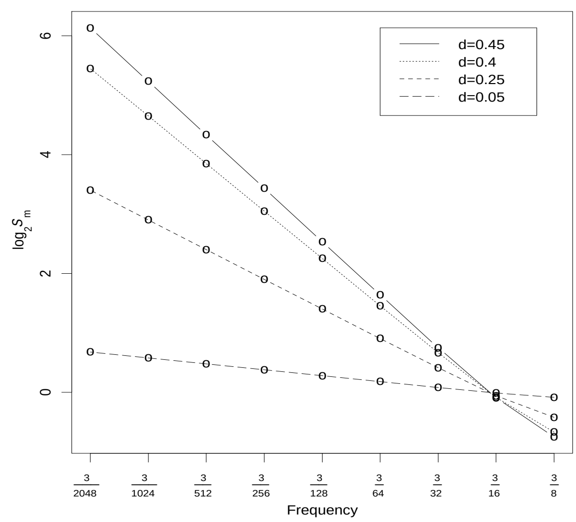


Figure 7.2: $\log_2 S_m$ against centre frequency for fdp's

Recall from Section 6.5 that when employing periodic boundary conditions, the sample mean and the 'smooth' coefficient of the discrete wavelet transform $(x_1^{N/2})$ are directly related. Other transform methods employing alternative boundary conditions do not have this property and may adversely affect the sample properties of the mean of simulations using the method we have described here, producing datasets with bias in the mean. It is for this reason that we choose periodic boundary conditions in the simulations shown in the following section.

7.4.2 Examples of generated series

Using the values of the bandpass variances given in Table 7.2 and the algorithm described in Table 7.1 we have simulated some fractionally differenced processes using a variety of wavelets and increasing values of d, and $\sigma_\epsilon^2 = 1$. Figures 7.3 to 7.6 show these simulated series, these figures can be compared to those simulated via Hosking's method shown in Figure 2.4 in Section 2.3. We can see that visually which wavelet you choose does not have a great impact on the synthesized series.

To assess whether the synthesized process has the correct structure, we will now examine both the actual autocovariance structure of the generated series, and the sample autocovariance sequences from series generated using the algorithm.

From Equations 7.3 and 7.4 we can calculate the covariance of the generated series. These covariances are shown for a variety of wavelets in Figures 7.8 to 7.10, the actual covariance is shown in Figure 7.7. We can see from these figures that, although there are the usual differences between the wavelets, they all capture the slow decay inherent in the true process.

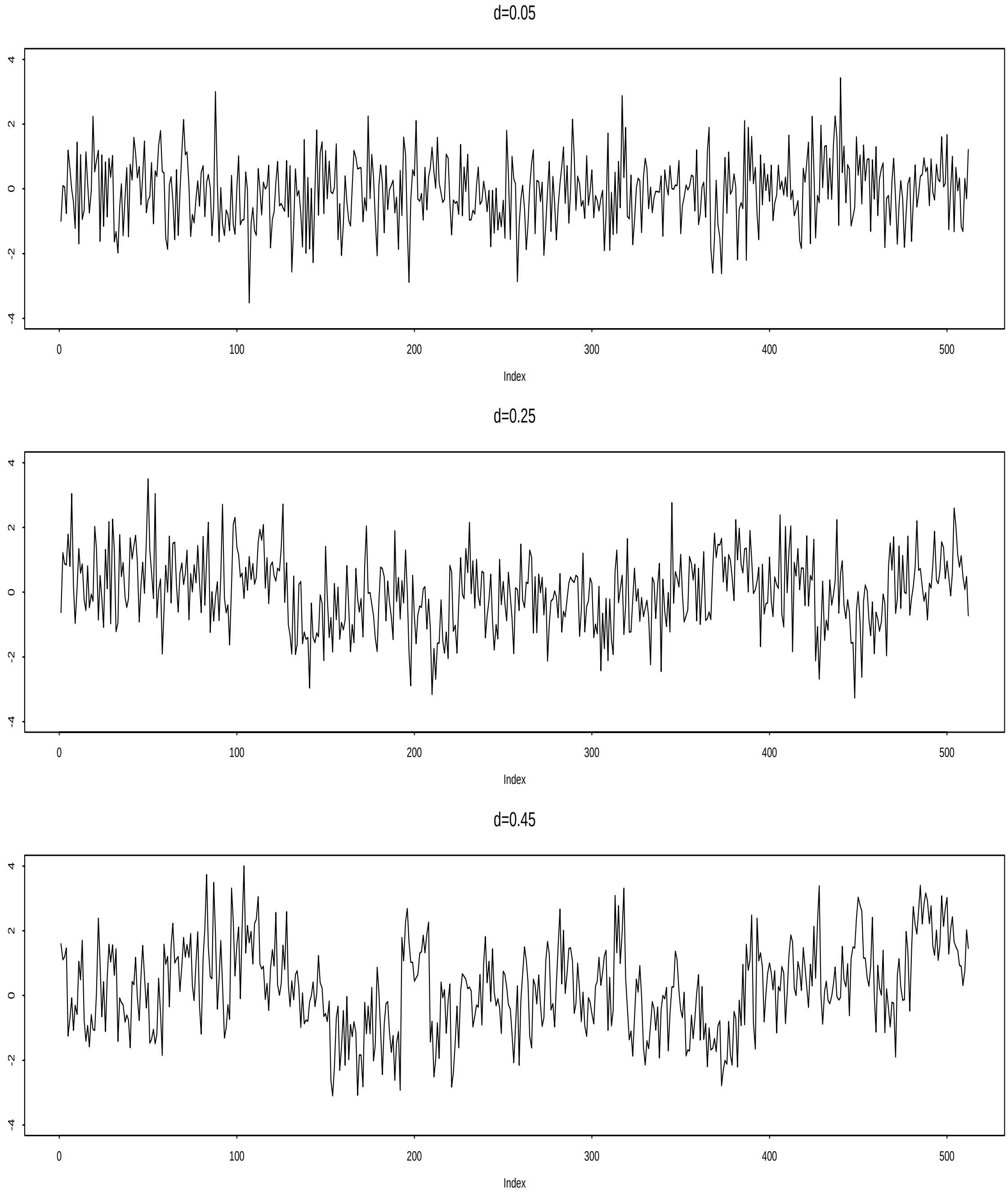


Figure 7.3: Simulated fractionally differenced processes using Haar

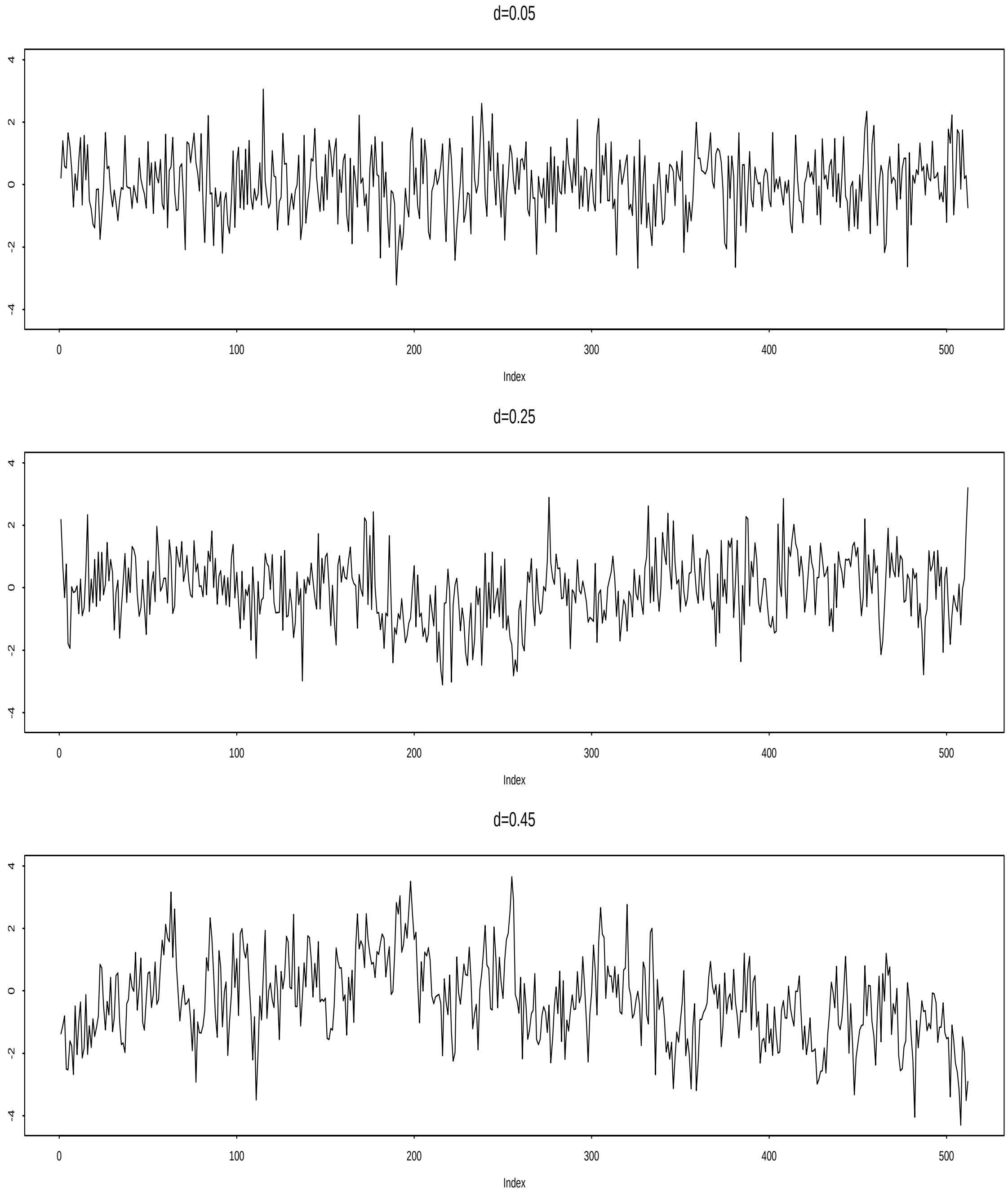


Figure 7.4: Simulated fractionally differenced processes using EP(4)

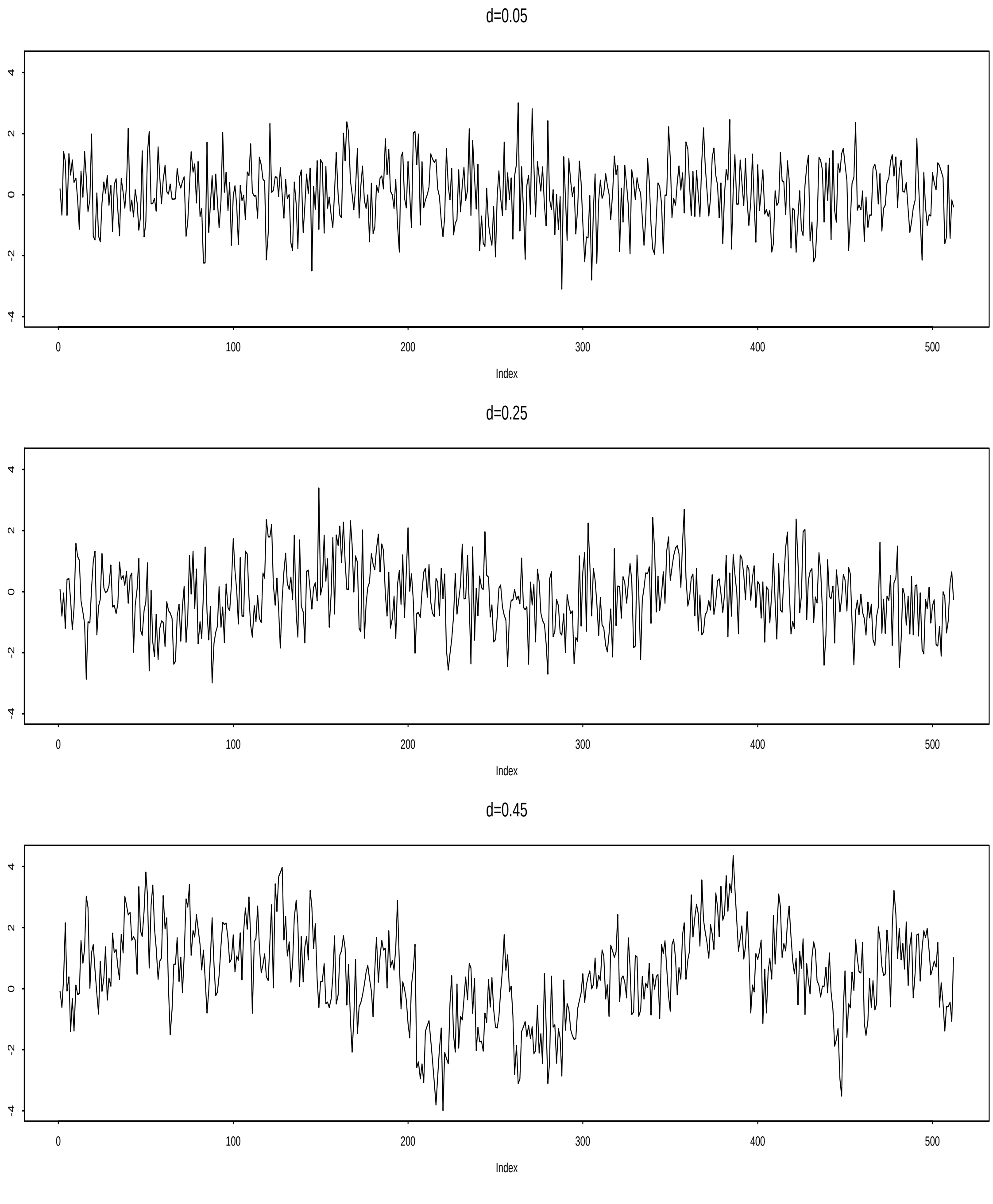


Figure 7.5: Simulated fractionally differenced processes using LA(8)

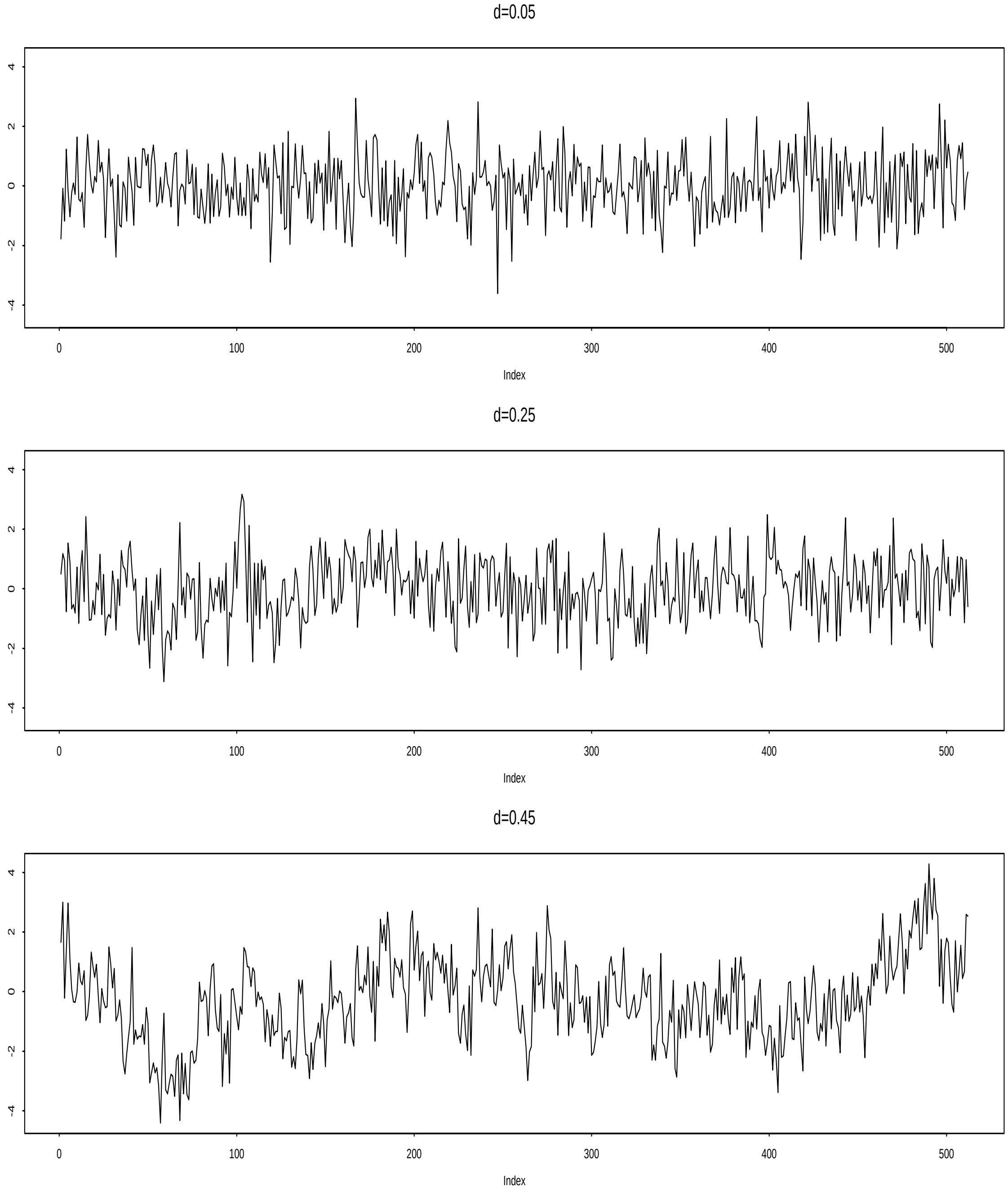


Figure 7.6: Simulated fractionally differenced processes using LA(20)

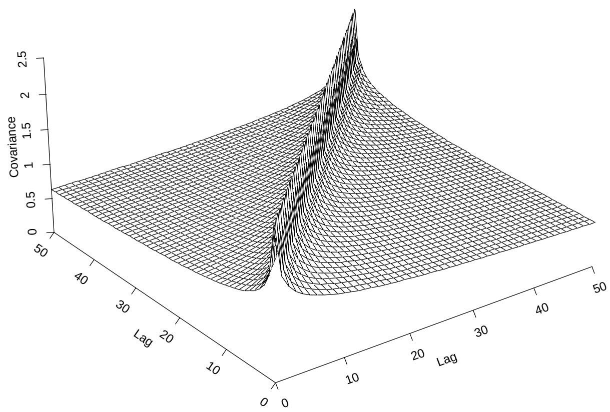


Figure 7.7: Covariance of an fdp(0.4)

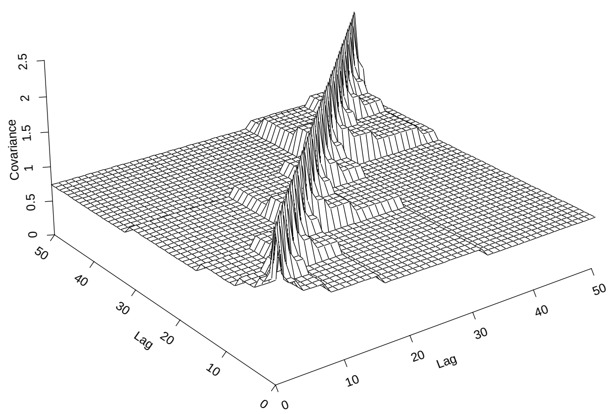


Figure 7.8: Covariance of generated series using Haar

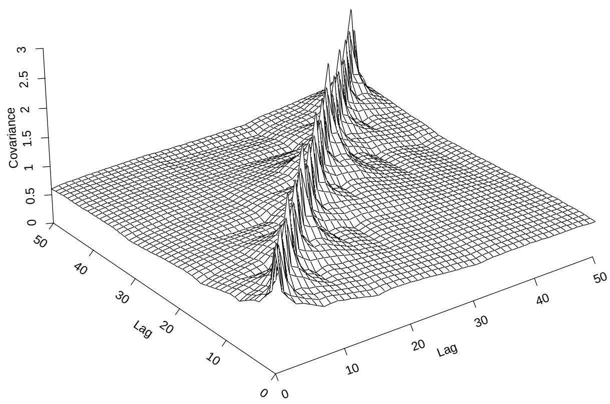


Figure 7.9: Covariance of generated series using EP(4)

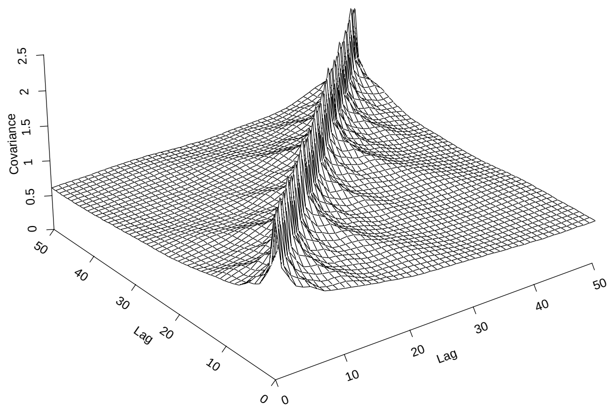


Figure 7.10: Covariance of generated series using LA(20)

We can also estimate the autocovariance sequence of the generated series by averaging along the diagonals of the matrices just described. These autocovariances (dotted line) are shown in Figures 7.11 to 7.13, verses the true autocovariance sequence (solid line), it is evident from these plots that in this average sense the wavelet method does extremely well in reproducing the correct covariance structure.

The real test of the method lies in how well any generated series reflect the

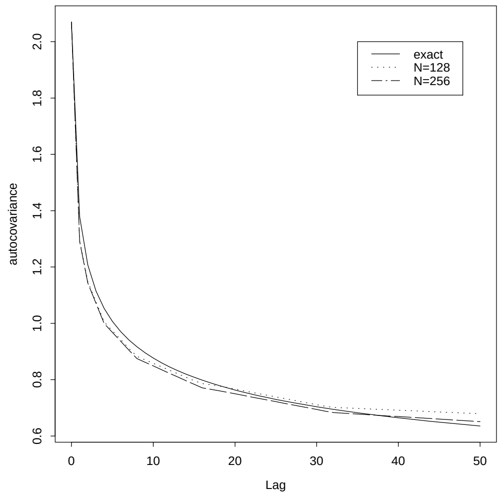


Figure 7.11: Autocovariance of generated series using Haar

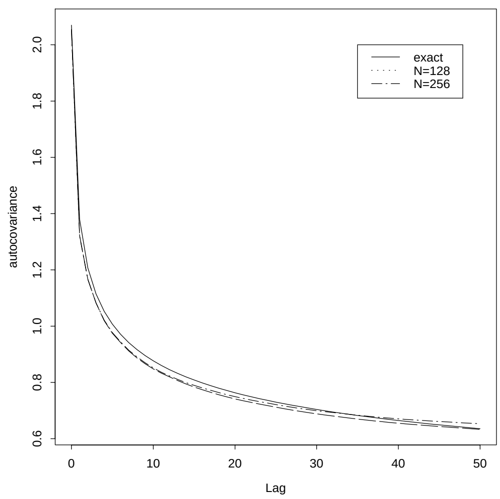


Figure 7.12: Autocovariance of generated series using EP(4)

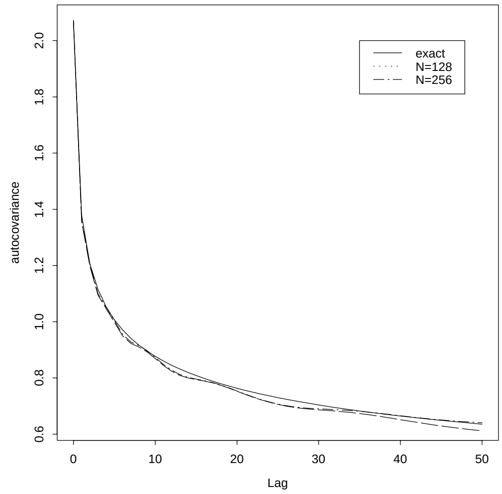


Figure 7.13: Autocovariance of generated series using LA(20)

true covariance structure. We generated 1000 series of differing lengths of an fdp(0.4) process using both the wavelet method and the Hosking scheme; for each generated series the sample autocovariance sequence was calculated, and these were then averaged over the 1000 series. These averaged sample autocovariance sequences and 5% and 95% quantiles are shown in Figures 7.14 to 7.17, showing that the performance of the wavelet and Hosking schemes are very similar, and both improve as the sample size increases from $N = 128$ to 256.

In conclusion, the wavelet synthesis method appears to provide an excellent method of simulating long memory processes, provided the bandpass variances are known.

7.5 Estimating using wavelets

We can exploit the fact that the coefficients of the discrete wavelet transform are approximately uncorrelated and have a known variance structure to estimate the parameters of the model. For zero mean Gaussian data, the discrete wavelet transform coefficients will also be zero mean Gaussian, thus

$$d_j^m \stackrel{d}{=} N(0, S_m) \quad \text{for} \begin{cases} m = 1, 2, 4, \ldots, N/2 \\ j = 1, 2, \ldots, N/2m, \end{cases}$$

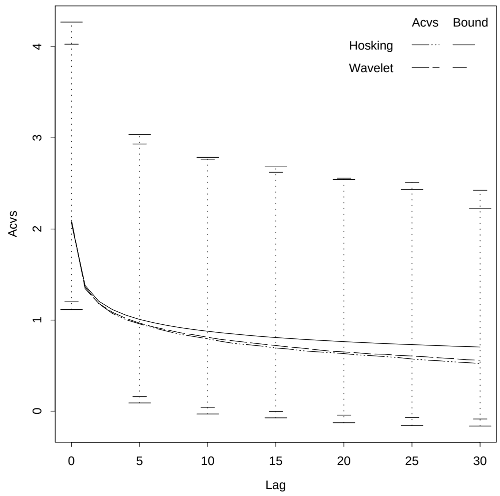


Figure 7.14: Average sample acvs of 1000 generated series of length 128 using EP(4)

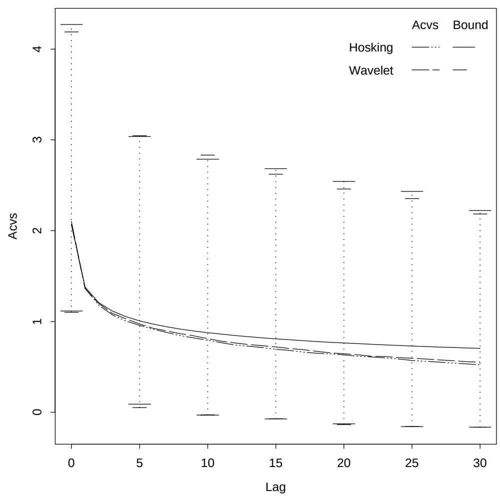


Figure 7.15: Average sample acvs of 1000 generated series of length 128 using LA(8)

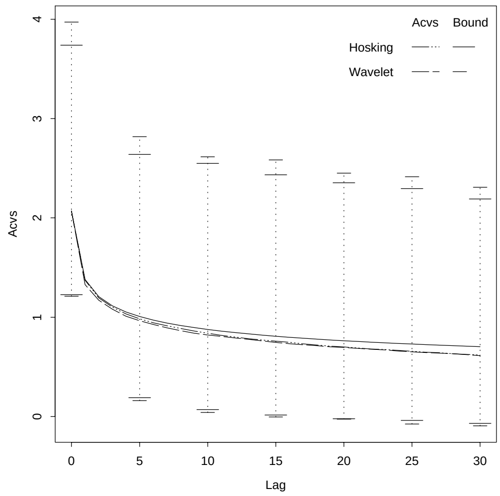


Figure 7.16: Average sample acvs of 1000 generated series of length 256 using EP(4)

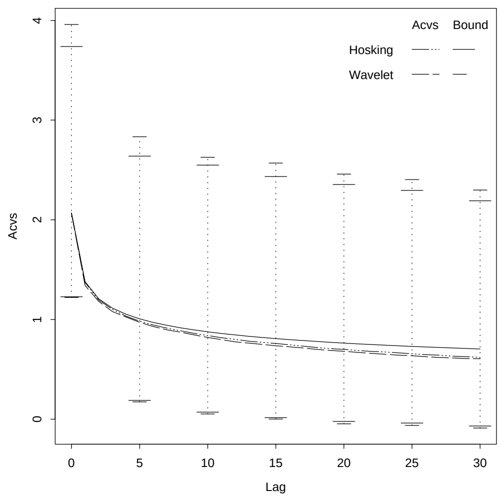


Figure 7.17: Average sample acvs of 1000 generated series of length 256 using LA(8)

where S_m is defined as in Equation 7.1 for purely $1/f$ processes or are derived from B_m as given in Table 7.2 for fractionally differenced processes. Also,

$$x_1^{N/2} \stackrel{d}{=} N(0, S_N),$$

where S_N is defined as in Equation 7.2 or derived from B_N in Table 7.2.

Letting $S_j = \sigma_\epsilon^2 S_j'$, we can then form the likelihood of the transform coefficients as

$$\frac{1}{\sqrt{2\pi\sigma_\epsilon^2 S_N'}} e^{-\frac{\left(x_1^{N/2}\right)^2}{2\sigma_\epsilon^2 S_N'}} \prod_m \prod_j \frac{1}{\sqrt{2\pi\sigma_\epsilon^2 S_m'}} e^{-\frac{\left(d_j^m\right)^2}{2\sigma_\epsilon^2 S_m'}}.$$

So our aim is to find values of the parameters d and σ_ϵ^2 which maximise the following function (derived from the log likelihood):

$$-\frac{N}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\log\{S_N'(d)\} - \sum_{m,j}\frac{1}{2}\log\{S_m'(d)\} - \frac{1}{2\sigma_\epsilon^2}\left\{\frac{\left(x_1^{N/2}\right)^2}{2S_N'(d)} + \sum_{m,j}\frac{\left(d_j^m\right)^2}{2S_m'(d)}\right\}.$$

Here we have made explicit the dependence of the S_j' on the parameter d.

Differentiating this function with respect to σ_ϵ^2 gives us the following estimate of σ_ϵ^2 which depends on the parameter d,

$$\widehat{\sigma_\epsilon^2(d)} = \frac{1}{N}\left\{\frac{\left(x_1^{N/2}\right)^2}{S_N'(d)} + \sum_{m,j}\frac{\left(d_j^m\right)^2}{S_m'(d)}\right\}. \tag{7.5}$$

Given an initial value for the parameter d, we can then obtain an estimator for σ_ϵ^2. We then form the concentrated log-likelihood by replacing σ_ϵ^2 by its estimator, and then maximise this concentrated log-likelihood to obtain an improved estimator for d.

The algorithm for obtaining the maximum likelihood estimators is shown in Table 7.3. This method of maximising the likelihood is similar in flavour to the method employed by Haslett and Raftery (1989), which employs a time-domain technique for estimating d and is discussed in Section 5.5.1.

For $1/f$ processes we can maximise the concentrated log-likelihood directly, but in the case of fractionally differenced processes, where we only have the band-pass variances numerically, the concentrated log-likelihood must also be maximised numerically.

1. Choose an initial estimate of d.

2. Calculate $S'_N(d)$ and $S'_m(d)$ as defined in Equation 7.1 and Equation 7.2.

3. Compute $\widehat{\sigma^2_\epsilon(d)}$ from Equation 7.5.

4. Compute the concentrated likelihood by replacing $\sigma^2_\epsilon(d)$ with its estimate.

5. Maximise this likelihood with respect to the parameter d, to obtain an improved estimator of d.

6. Goto **2**, and iterate until the desired accuracy of the estimator is obtained.

Table 7.3: Algorithm: Analysis of $1/f$-type processes via the discrete wavelet transform

7.5.1 Results for fractionally differenced processes

We have simulated 500 fractionally differenced datasets with $d = 0.25$ and $d = 0.4$, and $N = 128$ and $N = 256$, using Hosking's algorithm which was described in Table 2.1, Section 2.3, and have then estimated d using the maximum likelihood method outlined in the previous section.

The results of these simulations are shown in Table 7.4, where we display the mean, estimated bias, and estimated mean square error of the samples. The bias and mean square error (mse) are estimated in the same way as in Chapter 5, that is

$$\text{bias} = \frac{1}{500}\sum_{j=1}^{500}(\hat{d}_j - d)$$

and

$$\text{mean square error} = \frac{1}{500}\sum_{j=1}^{500}(\hat{d}_j - d)^2.$$

d		Haar		EP(4)		LA(8)	
		128	256	128	256	128	256
0.25	mean	0.2304	0.2297	0.2482	0.2421	0.2462	0.2470
	bias	-0.0211	-0.0203	-0.0088	-0.0080	-0.0039	-0.0030
	mse	0.0057	0.0030	0.0051	0.0026	0.0051	0.0025
0.4	mean	0.3681	0.3725	0.3853	0.3897	0.3911	0.3958
	bias	-0.0319	-0.0275	-0.0147	-0.0103	-0.0089	-0.0042
	mse	0.0050	0.0031	0.0040	0.0023	0.0037	0.0021

Table 7.4: Wavelet maximum likelihood simulation results

We see from these results that the wavelet method consistently produces estimates which are negatively biased, however the bias is very small, even for the larger value of d, and improves as the wavelet becomes more regular. The mean square error also reduces as the wavelet regularity increases, and does not appear to be affected by the severity of d. As we would expect, the results improve as the sample size is increased from length 128 to 256.

Figures 7.18 to 7.21 show boxplots of the results obtained from these simulations, which also show the results obtained using regression of the periodogram.

The horizontal line shows the true value of d. We can see that in all cases the wavelet method provides estimates with a lower variance than the periodogram method.

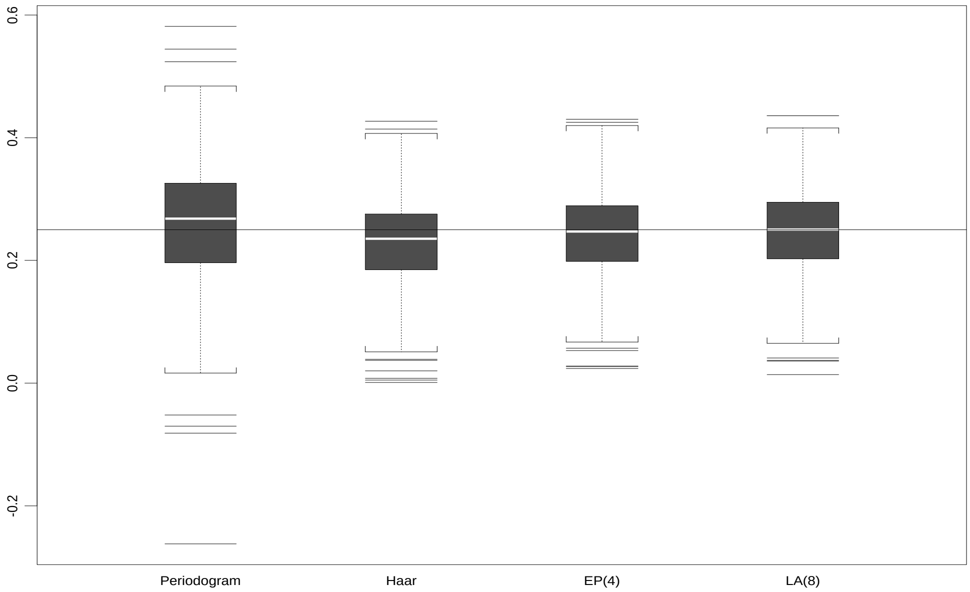


Figure 7.18: Estimates of d from 500 simulated datasets of length 128, $d = 0.25$

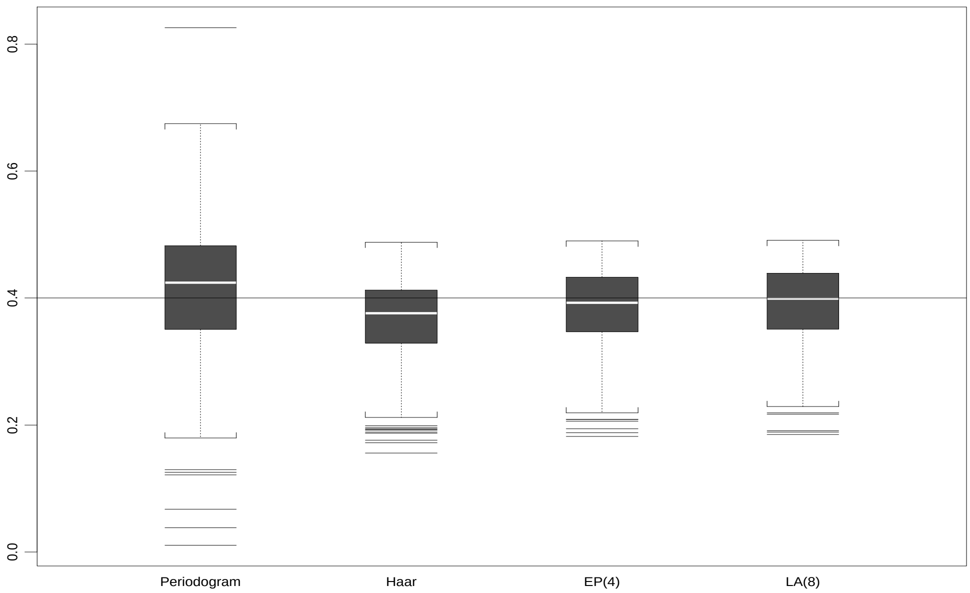


Figure 7.19: Estimates of d from 500 simulated datasets of length 128, $d = 0.4$

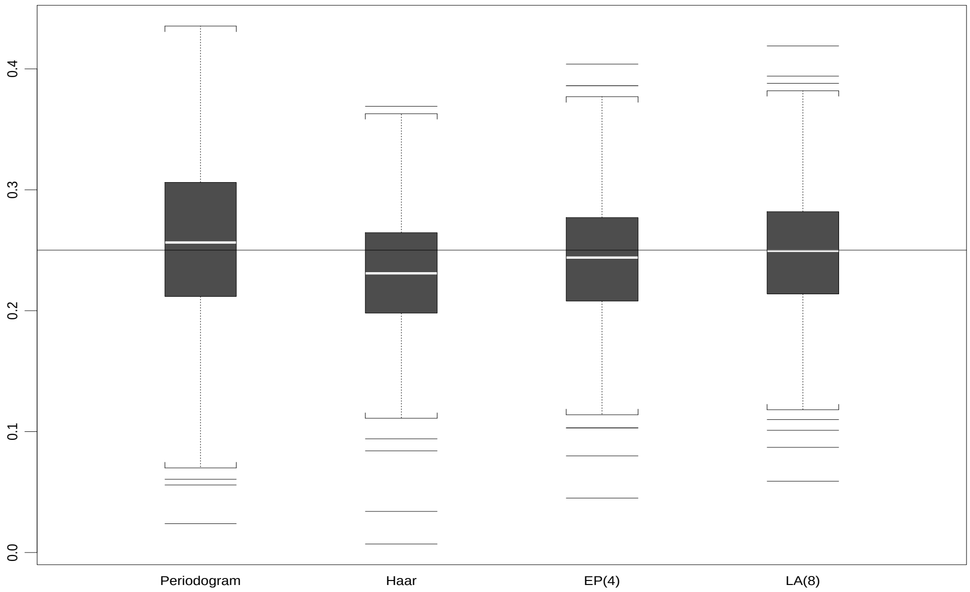


Figure 7.20: Estimates of d from 500 simulated datasets of length 256, $d = 0.25$

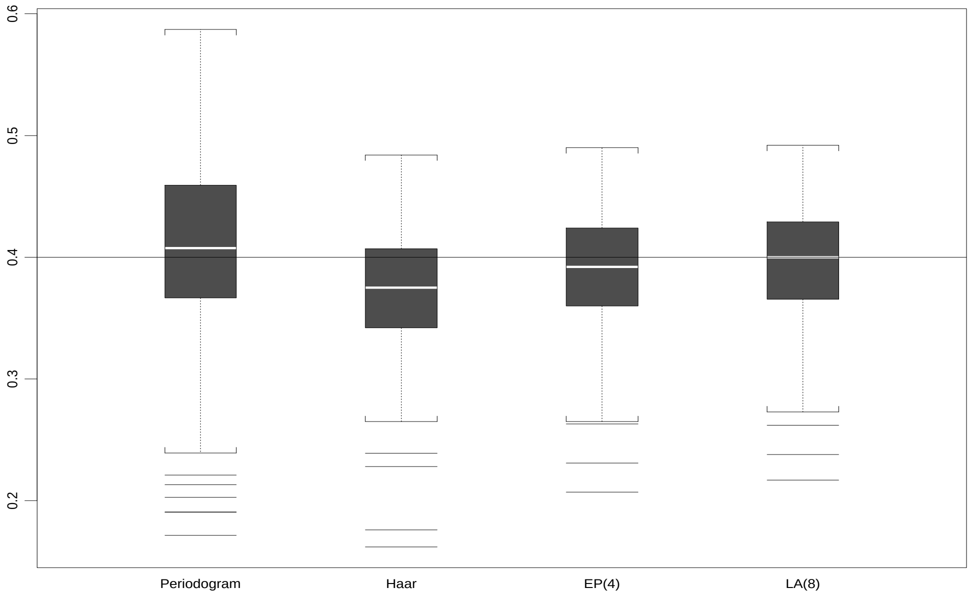


Figure 7.21: Estimates of d from 500 simulated datasets of length 256, $d = 0.4$

7.6 Comparison of the multitaper and wavelet methods

In this section we shall compare the estimates of d obtained using the regression method on the multitaper spectral estimates described in Chapter 5, and the maximum likelihood of the discrete wavelet transform outlined in this chapter. We deliberately used the same data in this chapter as was used in Chapter 5, so that such comparisons could be made.

We shall compare the 'best' estimates obtained from each of the methods, which are the 5 taper estimates and the LA(8) estimates. From Table 7.5 we can see that the wavelet method produces estimates with significantly lower bias than either the DPSS's or sinusoidal multitaper estimates. The multitaper method produces positively biased results, while the wavelet method produces negatively biased results. The mean square errors are comparable between the two methods, with the wavelet method producing results with smaller mean square error for the smaller sample size, though the improvement in the mean square error as the sample size increases is greater for the multitaper results.

Figures 7.22 and 7.23 show boxplots of these results, where we can clearly see the improvement in the bias when using the wavelet method over the multitaper method.

d		DPSS's		Sine tapers		LA(8)	
		128	256	128	256	128	256
0.25	mean	0.2717	0.2555	0.2738	0.2549	0.2462	0.2470
	bias	0.0217	0.0055	0.0238	0.0049	-0.0039	-0.0030
	mse	0.0076	0.0027	0.0068	0.0025	0.0051	0.0025
0.4	mean	0.4181	0.4092	0.4113	0.4063	0.3911	0.3958
	bias	0.0181	0.0092	0.0113	0.0063	-0.0089	-0.0042
	mse	0.0075	0.0018	0.0048	0.0019	0.0037	0.0021

Table 7.5: Comparison of multitaper regression and wavelet maximum likelihood

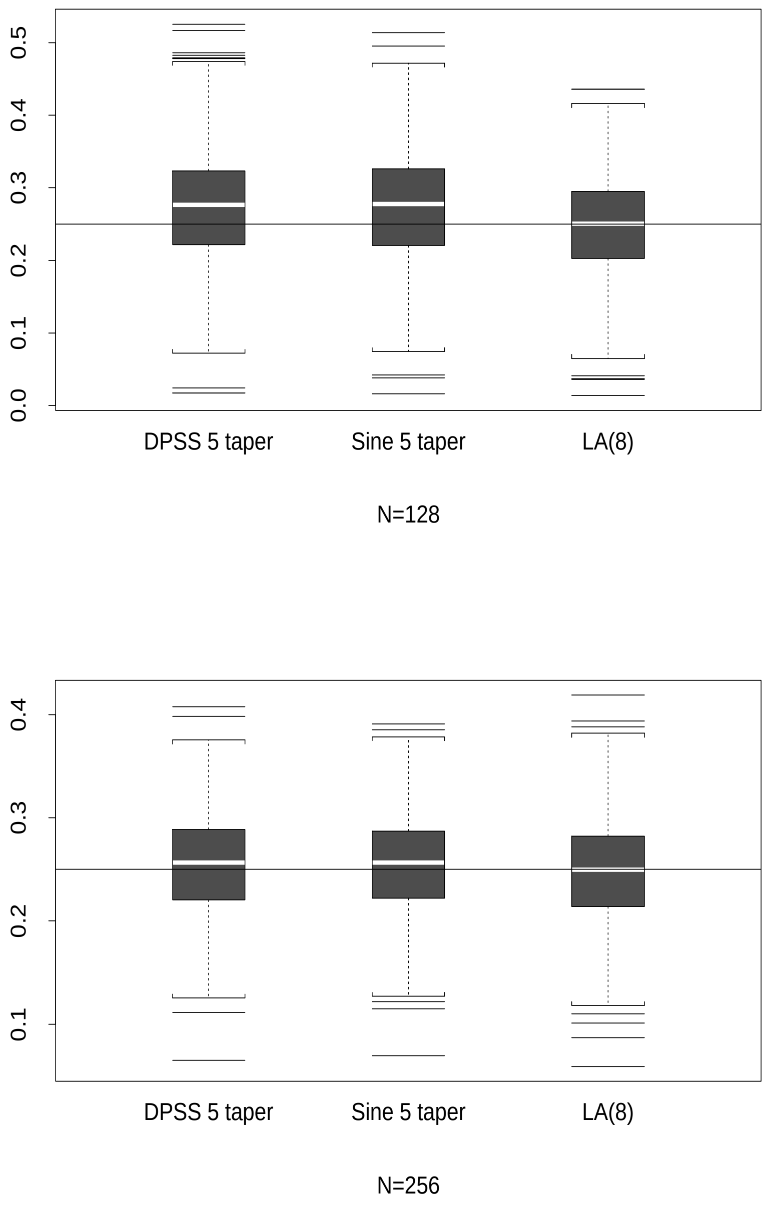
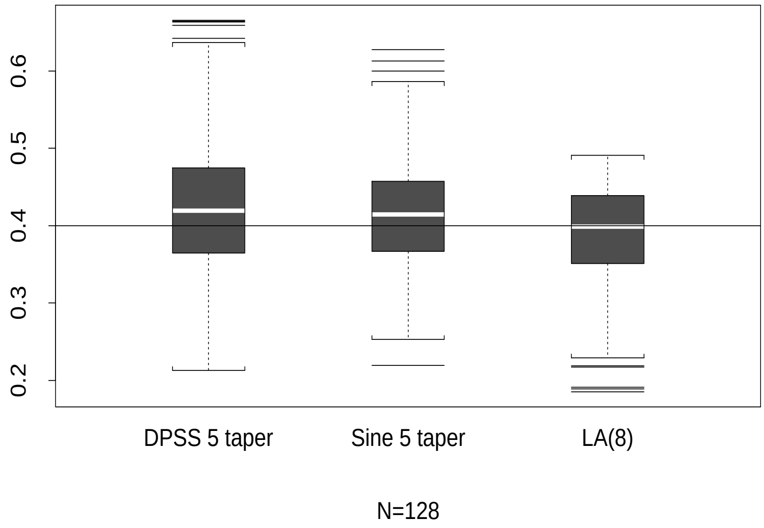


Figure 7.22: Comparison of multitaper regression and wavelet maximum likelihood, $d = 0.25$



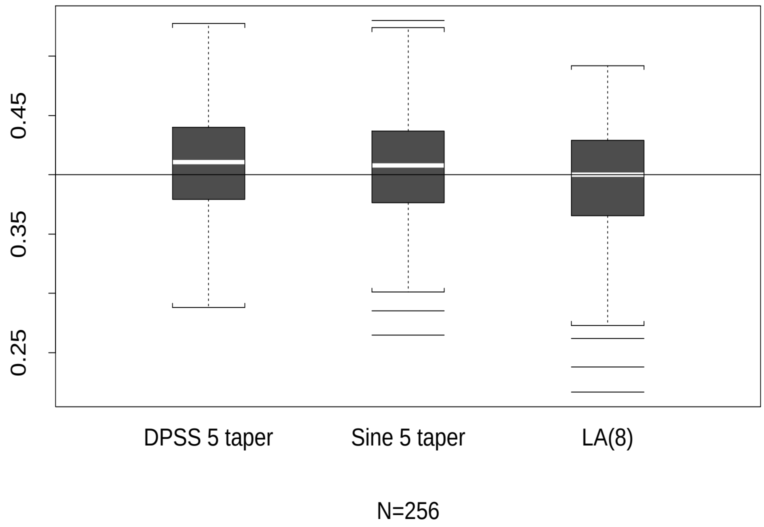


Figure 7.23: Comparison of multitaper regression and wavelet maximum likelihood, $d = 0.4$

Chapter 8

Conclusions

This thesis has looked at some new methods of analyzing time series which exhibit dependence over long periods of time. This dependence implies that the spectrum of the process has a large concentration of energy at low frequencies. We introduce a class of models capable of modeling long memory behaviour called fractionally differenced processes.

The first method we look at to analyze such processes is by using the multitaper spectral estimator, which reduces bias in the estimator due to leakage. We develop some general properties of multitaper spectral estimators; including the exact form of the variance of such an estimator when the autocovariance sequence of the data are known. We investigate the variance of multitaper estimators at low frequencies for fractionally differenced processes, showing that these variances can far exceed those derived using the ideal theoretical distributions of the estimator.

We develop a measure of bandwidth that can be used for multitaper spectral estimators, and use this in the method developed to estimate the parameter of a fractionally differenced white noise process – d. This method is based upon regression of the logarithm of the multitaper estimator, and performs very well when compared to some alternative methods suggested in the literature.

The next approach to analyzing long memory processes is by using the discrete wavelet transform. We develop synthesis and analysis algorithms which can be adapted to models with long memory behaviour – it is not restricted to particular models, but only depends upon the bandpass variances of the process of interest. We demonstrate the effectiveness of the wavelet synthesis algorithm by comparing the data to data synthesised using Hosking's exact method, and show that both methods produce data with almost identical properties. We use the wavelet analysis method to estimate d for data simulated using Hosking's method, and show that these estimates compare favourably with alternative previously published methods, and also produce less biased estimates than those constructed using the multitaper method.

We also investigate properties of the discrete wavelet transform, including the effect of boundary conditions employed in the transform on the sample mean of generated series, and the phase function of the transform which can effect interpretation of the discrete wavelet transform.

8.1 Further work

As both of the methods discussed in this thesis are relatively new there is still much work to be done, both in the general properties of the methods, and their application to long memory processes specifically.

Multitapering

A large outstanding problem in the area of multitaper spectral estimation is how to choose the number of tapers to use in a specific instance. One way of approaching this would be to use some form of cross-validation, perhaps employing the methods of Hurvich (1985) and Hurvich & Beltrao (1990), who develop cross-validation specifically for spectral estimators.

In this work we only investigated the simplest of fractionally differenced processes, the next step would be to apply multitapering to fuller models by introducing short memory behaviour into the model by way of autoregressive and moving average parameters. Another direction for research could be beyond the class of stationary invertible models to non-stationary models.

While multitapering solves the problem of bias in the spectral estimator, there are other methods such as pre-whitening which also deal with bias problems, the merits of these methods have not been investigated in full.

Wavelets

Here we have confined our attention solely to long memory processes and shown that the discrete wavelet transform of such a process is approximately uncorrelated both across and within scales. It would be interesting to investigate the correlation structure of other processes that exhibit both long and short memory, to see whether wavelets could provide a more general analysis or synthesis method. This is to be investigated in future work. We have not dealt with the question of stationarity of the synthesised process and more work needs to be completed in this direction.

Another interesting question arises from the impact of either or both trend and non-zero mean in the original data on the discrete wavelet transform.

The ultimate aim of these generalisations is to develop a theory which could provide a general time series analysis tool parallel to Fourier spectral analysis.

References

Abry P., Gonçalvès P. & Flandrin P. (1993). Wavelet based spectral analysis of $1/f$ processes. *IEEE Int. Conf. on Acoustics, Speech & Signal Processing*, Minneapolis, Minnesota, USA.

Agiakloglou C., Newbold P. & Wohar M. (1993). Bias in an estimator of the fractionally differenced parameter. *Journal of Time Series Analysis* **14**, 235–246.

Agiakloglou C. & Newbold P. (1994). Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis* **15**, 253–262.

Bartlett M.S. & Kendall D.G. (1946). The statisical analysis of variance-heterogeneity and the logarithmic transformation. *Supplement to the Journal of the Royal Statistical Society*, **8**, 128–138.

Bell B., Percival D.B. & Walden A.T. (1993). Calculating Thomson's spectral multitapers by inverse iteration. *J. Comp. Graph. Stat.* **2**, 119–130.

Beran J. (1989). A test of location for data with slowly decaying serial correlations. *Biometrika* **76**, 261–269.

Beran J. (1991). M-estimators of location for Gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association* **86**, 704–708.

Beran J. (1992a). A goodness-of-fit test for time series with long range dependence. *Journal of the Royal Statistical Society B* **54**, 749–760.

Beran J. (1992b). Statistical methods for data with long range dependence. *Statistical Sciences* **7**, 404–427.

Beran J. & Terrin N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *Journal of Time Series Analysis* **15**, 269–278.

Box G.E.P. & Jenkins G.M. (1976). *Time Series Analysis Forecasting and Control.* 2'nd ed., Holden Day, San Francisco.

Brillinger D.R. (1981). *Time Series: Data Analysis and Theory* (Expanded Edition). San Francisco: Holden-Day.

Bronez T.P. (1988). Spectral estimation of irregularly sampled multidimensional processes by generalised prolate spheroidal sequences. *IEEE Transactions on Acoustics, Speech and Signal Processing* **36**, 1862–1873.

Bronez T.P. (1992). On the performance advantage of mutitaper spectral analysis. *IEEE Transactions on Signal Processing* **40**, 2941–2946.

Burg J.P. (1967). Maximum entropy spectral analysis. In *Proceedings of the 37th Meeting of the Society of Exploration Geophysicists* (reprinted in Childers, 1981).

Chen G., Bovas A. & Peiris S. (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis* **15**, 473–487.

Childers D.G. editor (1981). *Modern Spectrum Analysis.* New York: IEEE Press.

Chui C.K. (1992). *An Introduction to Wavelets.* Boston: Academic Press.

Cohen A. & Daubechies I. (1993). Orthonormal bases of compactly supported wavelets III. Better frequency resolution. *SIAM Journal of Mathematical Analysis* **24**, 520–527.

Cohen A., Daubechies I. & Vial P. (1993a). Wavelets in the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis* **1**, 54–81.

Cohen A., Daubechies I., Jawerth B. & Vial P. (1993b). Multiresolution analysis, wavelets and fast algorithms on an interval. *Comptes Rendus Academie Sciences Paris, Series I* **316**, 417–421.

Dahlhaus R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics* **17**, 1749–1766.

Daubechies I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* **41**, 909–996.

Daubechies I. (1992a). *Ten Lectures on Wavelets.* Philadelphia: SIAM.

Daubechies I. (1992b). The wavelet transform: A method for time-frequency localization. In *Advances in Spectrum Analysis & Array Processing, Volume 1.* Simon Haydin ed. Prentice Hall.

Daubechies I. (1993). Orthonormal bases of compactly supported wavelets II. Variations on a theme. *SIAM Journal of Mathematical Analysis* **24**, 499–519.

Deeter J.E. & Boynton P.E. (1982). Techniques for the estimation of red power spectra. *The Astrophysical Journal* **261**, 337–350.

Deriche M. & Tewfik A.H. (1993). Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process. *IEEE Transactions on Signal Processing* **41**, 10, 2977–2989.

Fougère F. (1985). On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods. *Journal of Geophysical Research* **90**, 4355–4366.

Fox R. & Taqqu M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517–532

Geweke J. & Porter-Hudak S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 221–238.

Granger C.W. & Joyeux R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15–29.

Haslett J. & Raftery A.E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics* **38**, 1–50.

Hassler U. (1993). Regression of spectral estimator with fractionally integrated time series. *Journal of Time Series Analysis* **14**, 369–380.

Hosking J.R.M. (1981). Fractional differencing. *Biometrika* **68**, 165–176.

Hosking J.R.M. (1984). Modeling persistance in hydrological time series using fractional differencing. *Water Resources Research* **20**, 1898–1908.

Hurvich C.M. (1985). Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *Journal of the American Statistical Association* **80**, 933–940.

Hurvich C.M. & Beltrao K.I. (1990). Cross-validatory choice of a spectrum estimate and its connection with AIC. *Time Series Analysis* **11**, 121–137.

Hurvich C.M. & Beltrao K.I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *Journal of Time Series Analysis* **14**, 455–472.

Hurvich C.M. & Beltrao K.I. (1994). Automatic semiparametric estimation of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* **15**, 285–302.

Janacek G.J. (1982). Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis* **3**, 177–183.

Jenkins G.M. (1961). General considerations in the analysis of spectra. *Technometrics* **3**, 133–166.

Kaplan L.M. & Jay Kuo C.C. (1993). Fractal estimation from noisy data via discrete fractionsal Gaussian noise (DFGN) and the Haar basis. *IEEE Transactions on Signal Processing* **41**, 3554–3562.

Kashyap R.L. & Eom K-B. (1988). Estimation in long memory time series models. *Journal of Time Series Analysis* **9**, 35–41.

Li W.K. & McLeod A.I. (1986). Fractional time series modelling. *Biometrika* **73**, 217–221.

Kuo C., Lindberg C. & Thomson D.J. (1990). Coherence established between atmospheric carbon dioxide and global temperature. *Nature* **343**, 709–714.

Mallat S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intellegence* **11**, 674–693.

Mandelbrot B.B. & Van Ness J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437.

McHardy I. & Czerny B. (1987). Fractal X-Ray time variability and spectral invariance of the Seyfert galaxy NGC5506. *Letters to Nature* **325**, 696–698.

Meyer Y. (1993). *Wavelets: Algorithms & Applications.* Philadelphia: SIAM.

Nason G.P. & Silverman B.W. (1994). The discrete wavelet transform in S. *J. Comp. Graph. Stat.* **3**, 163–191.

Percival D.B. (1983). *Long Memory Processes.* PhD Thesis, University of Washington.

Percival D.B. & Guttorp P. (1994). Long memory processes, the Allan variance and wavelets. In *Wavelets in Geophysics*, E.Foufoula-Georgiou & P.Kumar eds., in press. New York: Academic Press.

Percival D.B. & Walden A.T. (1991). Multitaper spectral estimation for power law processes. *URSI Conference Paper.*

Percival D.B. & Walden A.T. (1993). *Spectral Analysis for Physical Applications.* Cambridge University Press.

Press W.H., Flannery B.P., Teukolsky S.A. & Vetterling W.T. (1992). *Numerical Recipes: The Art of Scientific Computing* (Second Edition). Cambridge University Press, Cambridge, England.

Priestley M.B. (1981). *Spectral Analysis and Time Series.* New York: Academic Press.

Ramsey F.L. (1974). Characterization of the partial autocorrelation function. *The Annals of Statistics*, **2**, 1296–1301.

Reisen V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *Journal of Time Series Analysis* **15**, 335–350.

Riedel K.S. & Sidorenko A. (1995). Mininum biased multitaper spectral estimation. *IEEE Transactions on Signal Processing*, to appear.

Rioul O. (1993). A discrete-time multiresolution theory. *IEEE Transactions on Signal Processing* **41**, 2591–2606.

Rioul O. & Duhamel P. (1992). Fast algorithms for discrete and continuous wavelet transforms. *IEEE Transactions on Information Theory* **38**, 569–586.

Rioul O. & Vetterli M. (1991). Wavelets and signal processing. *IEEE SP Magazine*, October 1991, 14–38.

Rutman J. (1978). Characterization of phase and frequency instabilities in precision frequency sources: Fifteen years of progress. *Proceedings of the IEEE* **66**, 1048–1075.

Samarov A. & Taqqu M.S. (1988). On the efficiency of the sample mean in long memory noise. *Journal of Time Series Analysis* **9**, 2, 191–200

Slepian D. (1978). Prolate spheroidal wave functions, Fourier analysis and uncertainty–V: the discrete case. *Bell Syst. Tech. J.* **57**, 1372–1430.

Smith R.L. (1992). Long range dependence and global warming. In *Statistics in the Environment*, V.Barnett & F.Turkman eds. Wiley, to appear.

Strang G. (1989). Wavelets and dilation equations: A brief introduction. *SIAM Review* **31**, 614–627.

Strang G. (1993). Wavelet transforms versus Fourier transforms. *Bulletin (New Series) of the American Mathematical Society* **28**, 288–305.

Strichartz R.S. (1993). How to make wavelets. *Am. Math. Mon.* **100**, 539–586.

Taqqu M.S. (1986). A bibliographic guide to self-similar processes and long-range dependence. *Dependence in Probability and Statistics*, E.Eberlein & M.S.Taqqu eds., Birkhäuser.

Tewfik A.H. & Kim M. (1992). Correlation structure of the discrete wavelet coefficients of fractional brownian motion. *IEEE Transactions on Information Theory* **38**, 569–586.

Tewfik A.H., Kim M. & Deriche M. (1993). Multiscale signal processing techniques: A review. In *Signal Processing and its Applications: Handbook of Statistics, Vol 10.* N.K.Bose and C.R.Rao, eds. Amsterdam, North-Holland.

Thomson D.J. (1977). Spectrum estimation techniques for characterization and development of WT4 waveguide - I. *Bell System Technical Journal* **56**, 1769–815.

Thomson D.J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* **70**, 1055–1096.

Vaidyanathan P.P. (1993). *Multirate Systems and Filter Banks.* Prentice Hall, Englewood Cliffs, New Jersey.

Vetterli M. & Herley C. (1992). Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing* **40**, 2207–2232.

Walden A.T. (1989). Accurate approximation of a 0'th order DPSS for filtering and data tapering. *Signal Processing* **18**, 341–348.

Walden A.T. (1994). Interpretation of geophysical borehole data via interpolation of fractionally differenced white noise. *Applied Statistics* **43**, 335–345.

Walden A.T. & Hosken J.W.J. (1985). An investigation into the spectral properties of primary reflection coefficients. *Geophysical Prospecting* **33**, 400–435.

Walden A.T., McCoy E.J. & Percival D.B. (1995). The effective bandwidth of a multitaper spectral estimator. *Biometrika*, to appear.

Walden A.T., McCoy E.J. & Percival D.B. (1994). The variance of multitaper spectrum estimates for real Gaussian processes. *IEEE Transactions on Signal Processing* **42**, 479–482.

Wornell G.W. (1990). A Karhunen-Loève-like expansion for $1/f$ processes via wavelets. *IEEE Transactions on Information Theory* **36**, 859–861.

Wornell G.W. (1991). Synthesis, analysis, and processing of fractal signals. *RLE Technical Report* **566**. Massachusetts Institute of Technology.

Wornell G.W. (1993). Wavelet-based representations for the $1/f$ family of fractal processes. *Proceedings of the IEEE, Special Section on Application of Fractals in Electrical Engineering.*

Wornell G.W. & Oppenheim A.V. (1992a). Wavelet-based representations for a class of self-similar signals with applications to fractal modulation. *IEEE Transactions on Information Theory* **38**, 785–800.

Wornell G.W. & Oppenheim A.V. (1992b). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing* **40**, 611–623.

Yaglom A.M. (1987). *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results.* New York: Springer-Verlag.

Yajima Y. (1988). On estimation of a regression model with long memory stationary errors. *The Annals of Statistics* **16**, 791–807.

Yajima Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *Journal of Time Series Analysis* **10**, 375–383.