Submitted to IEEE Trans. Sig. Proc., EDICS Cat. 3.1.1 or 3.6.1
Imperial College Statistics Section Technical Report TR-95-17

Multitaper Spectral Estimation of Power Law Processes

by

E.J.McCoy[1], A.T.Walden[1] & D.B.Percival[2]

ABSTRACT

In many branches of science, particularly astronomy and geophysics, power spectra of the form f^{β}, where β is a negative power-law exponent, are common. This form of spectrum is characterized by a sharp increase in the spectral density as the frequency f decreases towards zero. A power spectrum analysis method which has proven very powerful wherever the spectrum of interest is detailed and/or varies rapidly with a large dynamic range is the multitaper method. With multitaper spectral estimation a set of orthogonal tapers are applied to the time series and the resulting direct spectral estimators ('eigenspectra') are averaged, thus reducing the variance.

One class of processes with spectra of power-law type are fractionally differenced Gaussian white noise processes which are stationary and can model certain types of long range persistence. Spectral decay $f^{-\beta}$ can be modelled for $0 < \beta < 1$. Estimation of the spectral slope parameter by regression on multitaper spectral ordinates is examined for this class of processes. It is shown that multitapering, using sine or Slepian tapers, produces much better results than using the periodogram, and is attractive compared to other competing methods. The technique is applied to a geophysical estimation problem.

E.J. McCoy and A.T. Walden are with the Dept. of Mathematics, Imperial College of Science, Technology and Medicine, 180 Queen's Gate, London SW7 2BZ, U.K. D.B. Percival is with the Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195–5640, U.S.A. and Mathsoft, Inc., 1700 Westlake Ave. N, Seattle, WA 98109, U.S.A.

Correspondence to Prof. A.T. Walden, Dept. of Mathematics, Imperial College of Science, Technology and Medicine, 180 Queen's Gate, London SW7 2BZ, U.K. Phone (0)171 594 8524, E-mail: a.walden@ic.ac.uk

Revised version, 2/7/1997

I. INTRODUCTION

In many branches of science, particularly astronomy and geophysics, power spectra of the form f^{β}, where β is a negative power-law exponent, are common. The spectral density of such a process continues to increase as the frequency f decreases towards zero. The process is said to have long memory. When plotted on a log-log scale the power spectrum is linear over many decades of frequency. Some examples of such processes discussed in the scientific literature are (i) fluctuations in the Earth's rate of rotation [1], (ii) voltage fluctuations across cell membranes [2], (iii) the time series of impedances of rock layers derived from boreholes [3]–[5], and (iv) X-ray time variability of galaxies [6].

A common approach to estimating the exponent is to calculate the raw periodogram as an estimate of the spectrum, and then fit a straight line on a log-log scale [7],[8]. Here we discuss an approach which employs an alternative method of spectral estimation, that of multitapering. With multitaper spectral estimation a set of orthogonal tapers are applied to the time series and the resulting direct spectral estimators ('eigenspectra') are averaged, thus reducing the variance. In Section II we introduce the fractionally differenced Gaussian noise as a class of process with power-law spectral behaviour over many decades of frequency. In Section III we recall some facts on the periodogram, which are contrasted with multitaper spectral estimation in Section IV. The application of multitaper spectral estimators to the problem of estimating the exponent of a power-law process is discussed in Section V, and simulation results given. The multitaper approach is compared to autoregressive and likelihood approaches in Section VI, and in Section VII it is used to estimate the exponent in an important geophysical problem.

II. A STATIONARY LONG-MEMORY PROCESS

To compare different methods for estimating the power-law coefficient we need a convenient model for a stationary process with a power-law type spectrum. A suitable model is given by the fractionally differenced process (fdp). Let $\{X_t\}$ be a zero mean fdp process (e.g. [9]) with power spectral density function $S(f) = \sigma_\epsilon^2 (4 \sin^2 \pi f)^{-d}, \qquad |f| \le 1/2$. Then $S(f) \sim f^{-2d}$ as $f \rightarrow 0$. The process is stationary and long-memory for $0 < d < 1/2$. $\{X_t\}$ is related to a white noise process $\{\epsilon_t\}$ with mean zero and variance σ_ϵ^2 through $\nabla^d X_t = \epsilon_t$, where ∇ is the backward difference operator (i.e., $\nabla X_t \equiv (1-B)X_t = X_t - X_{t-1}$, where B is the backward shift operator defined by $BX_t \equiv X_{t-1}$) and the operator ∇^d is interpreted via the binomial series

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{1}{2} d(1-d)B^2 - \frac{1}{6}(1-d)(2-d)B^3 - \cdots$$

[9, Equation (2.1)]; i.e., the dth difference of $\{X_t\}$ is Gaussian white noise. This process can be obtained from a fractionally integrated ARMA(p,q) process, or ARFIMA(p,d,q), by setting $p = q = 0$ and so can be denoted as ARFIMA(0,d,0).

If we express power on a logarithmic scale,

$$\log S(f) = \log \sigma_\epsilon^2 - 2d \log(2|\sin \pi f|) \ , \tag{1}$$

then we see that the logarithm of the power spectrum is linearly related to the logarithm of $2|\sin \pi f|$. In fact since $\sin \pi f \approx \pi f$ for small f, it is also approximately linearly related to the logarithm of simple frequency f over decades of frequency close to zero. In Fig. 1 the power spectrum in dB is plotted (a) against $\log(f)$, illustrating the close-to-linear relationship, and (b) against $\log|2\sin(\pi f)|$, illustrating the exact linear relationship, for $d = 0.05, 0.25, 0.4$ and 0.45.

III. THE PERIODOGRAM

A. Basic Theory

Suppose we have a time series that is a realization of a portion $X_1, X_2, \ldots, X_N$ of the stationary process $\{X_t\}$ with zero mean, variance σ^2 and spectral density function $S(\cdot)$. The raw periodogram estimator of the spectrum is given by

$$\hat{S}^{(p)} \equiv \left| \sum_{t=1}^{N} \frac{1}{\sqrt{N}} X_t e^{-i2\pi f t} \right|^2 .$$

Now consider a stationary process with a spectral density $S(\cdot)$ that is continuous over the interval $[-1/2, 1/2]$ but that is not necessarily either purely random (white) or Gaussian distributed. Subject to the finiteness of high-order moments and the process having a small enough span of dependence (short-memory) to satisfy standard 'mixing' conditions (e.g., [10, p. 25], [11, p. 424]) then it can be proved that

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

asymptotically as $N \to \infty$. (Here the d above the equals sign means 'equal in distribution.') Note that for $0 < f < 1/2$, $\hat{S}^{(p)}(f)$ has only two degrees of freedom. Furthermore, for $0 \le f' < f \le 1/2$, $\hat{S}^{(p)}(f)/S(f)$ and $\hat{S}^{(p)}(f')/S(f)$ are asymptotically uncorrelated. We thus have

$$\begin{aligned} \mathrm{var}\left\{\hat{S}^{(p)}(f)\right\} &= S^2(f), \ \ 0 < f < 1/2, \\ \mathrm{cov}\left\{\hat{S}^{(p)}(f'), \hat{S}^{(p)}(f)\right\} &= 0, \ \ 0 \le f' < f \le 1/2, \end{aligned} \tag{3}$$

asymptotically as $N \to \infty$.

Now suppose the frequencies of interest are Fourier frequencies of the form $f = f_j = j/N$ for $0 < f_j < 1/2$. If the series is a Gaussian *purely random process*, then results (2) and (3) are exact, i.e., not only true asymptotically, [11, p.398]. However, if the series is Gaussian but non-white — as of interest here — then (2) and (3) are now only true asymptotically, [11, p. 425].

B. Periodogram Distribution: Long-memory Processes

Despite the fact that a long-memory fdp does not satisfy any mixing conditions [13], Yajima [32] proved that (2) still holds for any *fixed* frequency asymptotically as $N \to \infty$. Hurvich and Beltrao [33], (see also Robinson [34]), examined the distribution of $\hat{S}^{(p)}(f_j)/S(f_j)$ for $j = 1, ..., J$ for a fixed J as $N \to \infty$. They determine that for a Gaussian long-memory process the distribution is that of an unequally weighted linear combination of two independent χ_1^2 random variables. Here the Fourier frequencies *decrease* towards zero with increasing N, and are not fixed, as is the case in [32].

In order to ascertain the relevance of these large sample results in small sample situations, the following was done. Samples of length $N = 64$ from a fdp process with $d = 0.45$ were simulated using the algorithm in [12], and the periodogram calculated at selected Fourier frequencies. We concentrate on the Fourier frequencies $0 < f_j < 1/2$ given by $1/64$, $2/64$, $4/64$ and $16/64$. By repeating this process a thousand times, a sample distribution was built up of periodogram values at each of the selected Fourier frequencies. At each of these Fourier frequencies, the sample distribution was compared with the theoretical distribution of (2) using (a) quantile-quantile (q-q) plots, and (b) histograms.

To produce a q-q plot at any of the frequencies f_j, the periodogram values over the 1000 simulations are first ordered:

$$\hat{S}_1^{(p)}(f_j) \leq \hat{S}_2^{(p)}(f_j) \leq \cdots \leq \hat{S}_{1000}^{(p)}(f_j).$$

Then, for $l = 1, ..., 1000$ the corresponding theoretical quantile $q_l(f_j)$, of the chi-squared distribution scaled by $S(f_j)/2$, is calculated by solving

$$\text{Prob}\left\{\frac{S(f_j)\chi_2^2}{2} < q_l(f_j)\right\} = l/(N+1).$$

Example q-q plots are given in Fig. 2 for Fourier frequencies $1/64$, $2/64$, $4/64$ and $16/64$. Each plot is of $\hat{S}_l^{(p)}(f_j)$ versus $q_l(f_j)$ for $l = 1, ..., 1000$. The diagonal dotted lines mark the ideal 'perfect' fit. We see that the χ_2^2 approximation is good, (departures from the ideal fit not being severe), even for $N = 64$ and for frequencies near zero where the spectrum is changing rapidly.

Fig. 3 shows histograms of the $\hat{S}_1^{(p)}(f_j), ..., \hat{S}_{1000}^{(p)}(f_j)$ overlain by $S(f_j)\chi_2^2/2$ for the same Fourier frequencies as in Fig. 2. Again, there is no strong evidence refuting the use of (2) in practice.

These results are interesting considering that the fdp does not satisfy any mixing conditions [13].

C. Periodogram: Bias and Variance

The above simulation, and others, indicate that the theory developed for short-memory series can be extended to cover stationary fdp processes. The bias of the periodogram is very small for fdp processes. For example, calculations show that for $d = 0.45$ and $N = 64$,

$$|E\{\hat{S}^{(p)}(f_j) - S(f_j)\}| < 0.5\text{dB}$$

for $0 < f_j < 1/2$, with the largest deviation (0.49dB) occurring, not surprisingly, at f_1. However, the periodogram spectrum estimator is inconsistent since $\text{var}\{\hat{S}^{(p)}(f)\}$ does not decrease as $N \to \infty$. We now introduce multitapering as an approach for consistent variance reduction.

IV. TAPERING AND MULTITAPERING

A. Tapering

To taper the series we form the product $h_t X_t$ for each value of t, where $\{h_t\}$ is a suitable sequence of real-valued constants called a data taper. Let us define a direct spectral estimator (hence the superscript '(d)') as

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

Now,$E\{\hat{S}^{(d)}(f)\} = \int_{-1/2}^{1/2} \mathcal{H}(f - f')S(f')\, df'$, where $\mathcal{H}(f) = \left| \sum_{t=1}^{N} h_t e^{-i2\pi f t} \right|^2$. $\mathcal{H}(\cdot)$ is often called the *spectral window* of the direct spectral estimator $\hat{S}^{(d)}(\cdot)$. If $h_t = 1/\sqrt{N}$ for $1 \le t \le N$ (the so-called 'rectangular', 'default' or 'uniform' taper), $\mathcal{H}(\cdot)$ becomes Fejér's kernel $\mathcal{F}(\cdot)$ and $\hat{S}^{(d)}(\cdot)$ reverts to the simple periodogram. The key idea behind tapering is to select $\{h_t\}$ so that $\mathcal{H}(\cdot)$ has much smaller sidelobes than $\mathcal{F}(\cdot)$, thus alleviating the bias induced by sidelobe leakage. The direct spectral estimator is still inconsistent. As with the periodogram, for $0 < f < 1/2$, $\hat{S}^{(d)}(f)$ has only two degrees of freedom.

B. Variance Reduction And Consistency Using Multitapering

As its name implies, a *multitaper* spectral estimator [14],[15] utilizes several different data tapers, the kth of which is denoted $\{h_{t,k} : t = 1, ..., N\}$. These tapers are chosen to be orthonormal so that

$$\sum_{t=1}^{N} h_{t,j} h_{t,k} = \begin{cases} 0, & \text{if } j \ne k; \\ 1, & \text{otherwise.} \end{cases}$$

In its simplest formulation, this estimator is the average of K direct spectral estimators and hence takes the form

$$\hat{S}^{(mt)}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f) \tag{4}$$

with

$$\hat{S}_k^{(mt)}(f) \equiv \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi f t} \right|^2 ,$$

where $\{h_{t,k}\}$ is the data taper for the kth direct spectral estimator $\hat{S}_k^{(mt)}(\cdot)$. The estimator $\hat{S}_k^{(mt)}(\cdot)$ is also called the kth *eigenspectrum* but it is just a special case of the familiar direct spectral estimator $\hat{S}^{(d)}(\cdot)$. For each data taper, we can define an associated spectral window $\mathcal{H}_k(f) \equiv \left|\sum_{t=1}^{N} h_{t,k} \exp(-i2\pi f t)\right|^2$. Thus, $E\{\hat{S}^{(mt)}(f)\} = \int_{-1/2}^{1/2} \overline{\mathcal{H}}(f-f')S(f')\,df'$ and $\overline{\mathcal{H}}(f) \equiv \frac{1}{K}\sum_{k=0}^{K-1} \mathcal{H}_k(f)$. If the spectrum is not rapidly varying over the effective bandwidth of $\overline{\mathcal{H}}(f)$, then the orthogonality of the tapers implies that the eigenspectra are uncorrelated (see e.g., [16]) and, under the same conditions as for the results on the periodogram,

$$\hat{S}^{(mt)}(f) \stackrel{\mathrm{d}}{=} S(f)\chi^2_{2K}/2K, \;\; 0 < f < 1/2, \tag{5}$$

asymptotically as $N \to \infty$. Thus

$$\mathrm{var}\{\hat{S}^{(mt)}(f)\} = S^2(f)/K, \;\; 0 < f < 1/2.$$

The multitaper estimator is consistent provided K increases with N (see below). The spectrum estimates now have $2K$ degrees of freedom (2 for each of the K eigenspectra averaged). Provided several tapers are used, jackknifing can be used to estimate the variance of the spectral estimate [17]. Smoothed multitaper spectrum estimation is detailed in [36]. For extensive discussion and background information on multitapering, the reader is referred to the seminal paper [14] or to [15, Chapter 7].

The key then is to derive a set of K orthonormal sequences (tapers), each one of which does a satisfactory job of tapering. Until recently there has only really been one approach to designing the set of orthogonal tapers, and this was based on maximum sidelobe-bias protection. It produces the so-called *Slepian tapers*. However, another important aspect of multitapering is its variance reduction ability, without the need to lose resolution by smoothing across frequencies. Hence, if the spectrum is not changing too rapidly, so that sidelobe bias is not severe, alternative simpler designs are possible which will still lead to variance reduction compared to the periodogram. One such, the *Sine tapers*, was suggested by Riedel and Sidorenko [18].

C. Slepian Tapers

Let us denote the first of the orthogonal sequences of tapers by $\{v_{t,0},\ t = 1, ..., N\}$. This is the zeroth-order sequence. It is chosen such that its corresponding spectral window $\mathcal{V}_0(f) = \left|\sum_{t=1}^{N} v_{t,0} \exp(-i2\pi f t)\right|^2$ maximizes the concentration ratio over a design interval $[-W, W]$:

$$\frac{\int_{-W}^{W} \mathcal{V}_0 df}{\int_{-1/2}^{1/2} \mathcal{V}_0 df} = \lambda_0 \, .$$

The first-order sequence, $\{v_{t,1},\ t = 1, ..., N\}$ is also chosen to maximize this concentration ratio, but subject to being orthogonal to the zeroth-order sequence. The second-order sequence similarly maximizes the concentration ratio, but subject to being orthogonal to the previous two, etc. Maximization of the concentration ratio ensures that the sidelobes are minimized. λ_k must be close to unity for $\{v_{t,k}\}$ to be a decent taper, but this holds only for the first $K = 2NW - 1$ tapers, and hence at most K tapers should be used. If we fix W and then increase N, then K will increase and the spectrum estimator (4) is consistent. An algorithm, and instructions for obtaining code, for the calculation of the Slepian tapers is given in [19]. The average spectral window $\overline{\mathcal{V}}(f)$ is formed by averaging the first K of the $\mathcal{V}_k(f)$. Whereas all the tapers utilize the same nominal design interval $[-W, W]$, the effective bandwith of $\overline{\mathcal{V}}(f)$ increases with the number of tapers averaged [20], approaching $2W$ when all $2NW - 1$ tapers are used.

D. Sine Tapers

Riedel and Sidorenko [18] suggest a much simpler set of orthogonal tapers, the kth of which will be denoted $\{r_{t,k} : t = 1, ..., N\}$. The elements are given by

$$r_{t,k} = \left[\frac{2}{(N+1)}\right]^{1/2} \sin\left[\frac{(k+1)\pi t}{N+1}\right],$$

for orders $k = 0, ..., N-1$. The first term is a standardization factor to ensure that these tapers are orthonormal. Denote the spectral window of $\{r_{t,k}\}$ by $\mathcal{R}_k(f)$. Then the average spectral window $\overline{\mathcal{R}}(f)$ is formed by averaging the first K of the $\mathcal{R}_k(f)$. These sine tapers achieve a smaller local bias, i.e., the bias due to the smoothing by the main lobe of $\overline{\mathcal{R}}(f)$, than the Slepian tapers, at the expense of sidelobe suppression. Since these tapers have analytic expressions, it follows ([18]) that the kth sinusoidal taper has its spectral energy concentrated in the frequency bands

$$\frac{k}{2(N+1)} \leq |f| \leq \frac{k+2}{2(N+1)}, \quad k = 0, ..., N-1.$$

Use of K sine tapers yields a multitaper estimator with spectral window concentrated in the interval $[-W', W']$, where $W' \equiv (K+1)/\{2(N+1)\}$.

E. Multitaper Distribution For A Long-memory Example

The result (5) can be tested for multitapering in the same way as (1) for the periodogram, i.e., using q-q plots. It was noted earlier that a condition for (5) to hold, is that the spectrum should not rapidly vary over the effective bandwidth B of the average spectral window. For a frequency $0 < f < 1/2$, of interest this means that the spectrum should not change rapidly over the interval $[\max(f - B/2, 0), f + B/2]$. An expression for the effective bandwidth of the average spectral window was derived in [20]:

$$B\left\{\overline{\mathcal{H}}(\cdot)\right\} = \left\{ \sum_{\tau=-(N-1)}^{(N-1)} \left(\frac{1}{K} \sum_{k=0}^{K-1} \sum_{t=1}^{N-|\tau|} h_{t,k} h_{t+|\tau|,k} \right)^2 \right\}^{-1},$$

which is valid for Slepian and sine tapers. Quantile-quantile plots for multitaper (Slepian) spectral estimates ($N = 64$, $NW = 4$) of a fdp process with $d = 0.45$ are given in Fig. 4a (Fourier frequency of interest of $f = 2/64$), Fig. 4b ($f = 4/64$) and Fig. 4c ($f = 16/64$). In each figure the results using 1, 2, 3 and 4 tapers are given. With $N = 64, NW = 4$ the expression $f \pm B/2$ turns out as $f \pm 0.0222, f \pm 0.0322, f \pm 0.0395, f \pm 0.0457$, for 1, 2, 3 and 4 tapers, respectively. For the Fourier frequency $f = 0.03125 = 2/64$ (Fig. 4a) the interval $[\max(f - B/2, 0), f + B/2]$ is $[0.0091, 0.0535]$, $[0.0009, 0.0634]$, $[0, 0.0708]$, $[0, 0.0769]$ for 1, 2, 3 and 4 tapers, respectively. Such intervals include the area in which the spectrum is changing rapidly — for example there is a change of 17dB in spectrum power between $f = 0.001$ and $f = 0.08$ — and hence we are not surprised that (5) is an unsuitable distribution at this Fourier frequency. However, it is perhaps surprising just how poor it is! At the Fourier frequency $f = 0.0625$, (Fig. 4b) the values of $[\max(f - B/2, 0), f + B/2]$ are $[0.0403, 0.0847]$, $[0.0303, 0.0946]$, $[0.0230, 0.1020]$, $[0.0168, 0.1082]$ for 1, 2, 3 and 4 tapers, respectively. In these intervals the spectrum is changing much less rapidly — there is a change of 3.5dB in spectrum power between $f = 0.04$ and $f = 0.1$ —and this is reflected in the q-q plots which indicate that here the distribution (5) is not too poor an approximation. At frequency $f = 0.25$ all the intervals $[\max(f-B/2, 0), f+B/2]$ are well away from zero, so that the spectrum is not rapidly changing and the approximation (5) is very good (Fig. 4c). In summary then, we expect (5) to break down badly when $f \leq B/2$. As seen for one and two tapers in Fig. 4a, (5) can also perform poorly if f only just exceeds $B/2$ (by about 0.009 and 0.0009 for one and two tapers, respectively, in the example).

V. REGRESSION ON MULTITAPER ORDINATES

A. Regression Model

From (1), at Fourier frequencies f_j:

$$\log S(f_j) = \log \sigma_\epsilon^2 - 2d \log(2|\sin \pi f_j|)\ , \tag{6}$$

Adding $\log\{\hat{S}^{(mt)}(f_j)\}$ to both sides of (6),

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

So, we are in a position to estimate d using the linear regression $Y_j = \alpha x_j + \varepsilon_j,\ \ j = 1, \ldots, M$, where $M = \lfloor (N-1)/2 \rfloor$, $Y_j = \log\{\hat{S}^{(mt)}(f_j)\}$, $x_j = 2\log\{2|\sin \pi f_j|\}$, $\alpha = -d$, and $\varepsilon_j = \log\{\sigma_\epsilon^2 \hat{S}^{(mt)}(f_j)/S(f_j)\}$. Given (5), the distribution of the $\{\varepsilon_j\}$ is

$$\varepsilon_j \stackrel{\mathrm{d}}{=} \log\left\{ \frac{\sigma_\epsilon^2}{2K} \chi_{2K}^2 \right\}.$$

Then $E\{\varepsilon_j\} = \log \sigma_\epsilon^2 + \psi(K) - \log(K)$ and $\mathrm{var}\{\varepsilon_j\} = \psi'(K)$, where ψ and ψ' are, respectively, the digamma and trigamma functions. If we rewrite the regression as

$$Y_j = c + \alpha x_j + \eta_j, \quad j = 1, \ldots, M, \tag{7}$$

where $c = \log \sigma_\epsilon^2 + \psi(K) - \log(K)$ and $\eta_j = \varepsilon_j - [\log \sigma_\epsilon^2 + \psi(K) - \log(K)]$, then $E\{\eta_j\} = 0$ and $\mathrm{var}\{\eta_j\} = \sigma_\eta^2 = \psi'(K)$.

The distribution of the 'error term' $\{\eta_j\}$ — which is a $\log \chi_{2K}^2$ random variable — becomes rapidly more Gaussian as the number of tapers (and hence the degrees of freedom $2K$) increases; the results in [21] suggest that five or more tapers should provide a reasonable approximation to Gaussianity. In the simulations which follow we use $K = 5$ in order that the approximate Gaussianity is valid.

B. Covariance Structure

The η_j sequence in the regression equation (7) is not a sequence of independent random variables since the associated spectral estimator is not uncorrelated at the Fourier frequencies. In order to make any progress in attempting to quantify the correlation structure of the η_j we assume that we are dealing with a spectrum which is locally slowly varying so that $S(f_j) \approx S(f_{j+\tau})$ for frequencies separated by less than the bandwidth of the estimator. (For a power law process, this assumption will of course be very bad at low frequencies.) Let us define $u_j = \hat{S}^{(mt)}(f_j)/S(f_j)$. Then,

$$\begin{aligned} s_{\eta,\tau} \equiv \mathrm{cov}\{\eta_j, \eta_{j+\tau}\} &= \mathrm{cov}\{\log(u_j), \log(u_{j+\tau})\} \\ &= \mathrm{cov}\left\{\log\left[\frac{\hat{S}^{(mt)}(f_j)}{S(f_j)}\right], \log\left[\frac{\hat{S}^{(mt)}(f_{j+\tau})}{S(f_{j+\tau})}\right]\right\}. \end{aligned}$$

Now, since η_j is approximately normal, for five or more tapers, $s_{\eta,\tau} \approx \log(1 + s_{u,\tau})$, [25, Section 3.7], a result which follows directly from the bivariate normal distribution of $(\eta_j, \eta_{j+\tau})$. To evaluate $s_{u,\tau}$ we note that

$$\begin{aligned} s_{u,\tau} &= \mathrm{cov}\left\{\frac{\hat{S}^{(mt)}(f_j)}{S(f_j)}, \frac{\hat{S}^{(mt)}(f_{j+\tau})}{S(f_{j+\tau})}\right\} \\ &= \mathrm{cov}\left\{\frac{\sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f_j)}{KS(f_j)}, \frac{\sum_{l=0}^{K-1} \hat{S}_l^{(mt)}(f_{j+\tau})}{KS(f_{j+\tau})}\right\} \\ &= \frac{1}{K^2} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \mathrm{cov}\left\{\frac{\hat{S}_k^{(mt)}(f_j)}{S(f_j)}, \frac{\hat{S}_l^{(mt)}(f_{j+\tau})}{S(f_{j+\tau})}\right\} \\ &\approx \frac{1}{K^2 S^2(f_j)} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \mathrm{cov}\left\{\hat{S}_k^{(mt)}(f_j), \hat{S}_l^{(mt)}(f_{j+\tau})\right\} \end{aligned} \tag{8}$$

under the assumption of a locally slowly varying spectrum. Thomson [14, p. 1069] showed that under the same assumption

$$\mathrm{cov}\left\{\hat{S}_k^{(mt)}(f_j), \hat{S}_l^{(mt)}(f_{j+\tau})\right\} \approx S^2(f_j) \left|\sum_{t=1}^{N} h_{t,k} h_{t,l} e^{i2\pi\tau t/N}\right|^2.$$

Thus, for a locally slowly varying spectrum,

$$s_{\eta,\tau} \approx \log\left(1 + \frac{1}{K^2}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\left|\sum_{t=1}^{N} h_{t,k}h_{t,l}e^{i2\pi\tau t/N}\right|^2\right),$$

which provides an approximation to the covariance structure of the 'noise' in (7).

As pointed out previously, the spectral window of the multitaper spectrum estimator using sine tapers is concentrated in the interval $[-W', W']$ where $W' = (K+1)/2(N+1)$. The bandwidth of the estimator is hence $2W' = (K+1)/(N+1)$. This means that the multitaper spectrum estimator at Fourier frequencies f_j and f_{j+K+1} should be essentially uncorrelated. As shown in [35], a simple, accurate and convenient model for $\{s_{\eta,\tau}\}$ for sine tapers is

$$s_{\eta,\tau} = \begin{cases} \sigma_\eta^2\left(1 - \frac{|\tau|}{K+1}\right), & \text{if } |\tau| \leq (K+1); \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Having thus a good model for the covariance structure of the noise in the linear regression (7), we are in a position to consider ordinary least-squares estimation of d, and the consequent variance of the estimator, and also weighted least-squares estimation of d, with the weighting determined by $\{s_{\eta,\tau}\}$.

C. Ordinary (Unweighted) Least Squares

We can write the M equations of (7) in the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{N} \tag{10}$$

where $\mathbf{Y}$ is the $M \times 1$ column vector of Y_j's, $\mathbf{X}$ is the $M \times 2$ matrix with first column all 1's and second column consisting of the x_j's, $\boldsymbol{\theta}$ is the 2×1 column vector of parameters c and α, and $\mathbf{N}$ is the $M \times 1$ column vector of the η_j's.

As is well-known [11, p. 592], the least-squares estimator corresponding to (10) is

$$\hat{\boldsymbol{\theta}} = \left[\mathbf{X}^T\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{Y} = \boldsymbol{\theta} + \left[\mathbf{X}^T\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{N}.$$

Since $\mathbf{N}$ has mean zero, $\hat{\boldsymbol{\theta}}$ is an unbiased estimator, and its mean square error is given by

$$E\left\{\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right]\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right]^T\right\} = \left[\mathbf{X}^T\mathbf{X}\right]^{-1}\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X}\left[\mathbf{X}^T\mathbf{X}\right]^{-1}, \tag{11}$$

where $\boldsymbol{\Sigma}$ is the Toeplitz covariance matrix of $\mathbf{N}$, which can be modelled using (9) when several sine tapers are used.

Estimating $d = -\alpha$ by least-squares regression leads to the estimator

$$\hat{d}^{(mt)} = -\hat{\alpha}^{(mt)} = -\frac{\sum_{j=1}^{M}(x_j - \bar{x})Y_j}{\sum_{j=1}^{M}(x_j - \bar{x})^2}.$$

From the appropriate entry of the 2×2 covariance matrix in (11), $\mathrm{var}\{\hat{d}^{(mt)}\}$ is given by

$$\psi'(K)\frac{\sum_{l=0}^{N-1}(x_{l+1}-\bar{x})\sum_{j=\max(-l,-K)}^{\min(K,N-l-1)}(x_{j+l+1}-\bar{x})\Delta_j}{\left[\sum_{i=1}^{M}(x_i-\bar{x})^2\right]^2},$$

where

$$\Delta_j \equiv \begin{cases} 1-\frac{|j|}{K+1}, & j=0,1,\ldots,K; \\ 0, & \text{otherwise.} \end{cases}$$

Standard deviations of $\hat{d}^{(mt)}$ are compared for $K = 4, 5$ and 6 sine tapers in Table 1, along with the standard deviation of the estimator $\hat{d}^{(p)}$ obtained from regression based on the periodogram, which has variance given by

$$\mathrm{var}\{\hat{d}^{(p)}\} = \frac{\pi^2}{6\sum_{i=1}^{M}(x_i-\bar{x})^2}.$$

(As pointed out in [22, p.532] this variance is discussed in [7] only for $d < 0$, the short-memory case, but their simulation results support its validity for $0 < d < 0.5$, the long-memory case). From Table 1 we note for example for $K = 5$ that the standard deviation of $\hat{d}^{(mt)}$ corresponds to a reduction of about 20% over that obtained from the periodogram.

The theoretical standard deviations of $\hat{d}^{(mt)}$ presented in Table 1 have been derived under various approximations, most notably perhaps, that the spectrum is locally slowly varying. How do empirical results compare? Five hundred realizations of fdp processes with d of 0.25 and 0.4 and lengths $N = 128$ and 256 were simulated using the algorithm in [12], taking $\sigma_\epsilon^2 = 1$. The parameter d was estimated (i) by regression on all the periodogram ordinates, and (ii) by regression on multitaper ordinates excluding those that are closer to zero frequency than $B/2$. (Recall that for ordinates closer to zero frequency than $B/2$ we know the distribution (5) is inappropriate).

Table 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: the empirical bias is defined as the average of the 500 values of $\hat{d}^{(mt)}$, minus the true value of d, while the standard error, (estimated standard deviation), is defined in the usual way. The periodogram estimates of d tend to be less biased than the multitaper estimates, but have a considerably larger standard error, the latter agreeing closely with the theoretical standard deviations of the estimators given in Table 1. Our periodogram results produce estimators with lower bias and mse than those published in either [23] or [24]; this is due to the fact that these authors do not include *higher* frequency ordinates in the regression, since they are studying models incorporating short memory, whose effect they wish to minimize when estimating the long-memory parameter d; see also [37],[38]. The sinusoidal tapers perform marginally better than the discrete prolate spheroidal sequences in these simulations, but when $N = 256$ results are virtually identical. For five sine tapers and $d = 0.25$ there is good agreement between the standard deviations in Table 1 (0.0810 and 0.0567 for $N = 128$ and 256 respectively), and the empirical standard errors in Table 2 (0.0790 and 0.0496, respectively). For $d = 0.4$ the agreement is not so good; for this model the spectrum is changing more rapidly.

Another way to compare the estimates derived from the different methods is via boxplots. 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×(inter-quartile range). Points that are very extreme are plotted as isolated horizontal bars. Boxplots comparing estimates of d obtained using the periodogram and multitapering (sine tapers) for $d = 0.4$ and $N = 128$ are shown in Fig. 5. 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, the estimate of d improves. However, the bias in the estimate also appears to positively increase as the number of tapers increases, due to the increase in bandwidth of the mean spectral window.

D. Weighted Least Squares

The weighted least-squares estimator corresponding to (10) is [11, p. 592],

$$\hat{\boldsymbol{\theta}}_W = \left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y} = \boldsymbol{\theta} + \left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{N}.$$

Hence, $\hat{\boldsymbol{\theta}}_W$ is an unbiased estimator, and its mean square error is given by

$$E\left\{\left[\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}\right]\left[\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}\right]^T\right\} = \left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]^{-1}. \tag{12}$$

The variance of the weighted least squares estimate $\hat{d}_W^{(mt)}$ of d follows from the appropriate entry of the 2×2 covariance matrix in (12). Theoretical standard deviations of $\hat{d}_W^{(mt)}$ for 5 sine tapers and $N = 128$ and 256 are given in Table 1.

Fig. 6 shows comparisons, in boxplot form, of estimates of d, calculated using the raw periodogram, the ordinary (unweighted) least squares approach, and the weighted least squares approach, for $N = 128$ and 256, and $d = 0.4$ using 5 sine tapers. The simulation results for weighted least squares and $d = 0.25$ and 0.4 are also included in Table 2. Comparing the unweighted and weighted schemes overall, the results of weighting are marginally beneficial, with a slight reduction in bias, and some reduction in scatter. However, the weighting scheme is heavier computationally due to forming-up and inverting the covariance matrix. For weighted least squares using five sine tapers and $d = 0.4$ and $N = 128$ there is good agreement between the theoretical standard deviation in Table 1 of 0.0626 and the empirical standard error in Table 2 of 0.0609; for $d = 0.25$ and $N = 256$ there is almost perfect agreement between the theoretical standard deviation in Table 1 of 0.0473 and the empirical standard error of 0.0472. In what follows, we refer only to the unweighted scheme.

VI. COMPARISON WITH OTHER METHODS

A. Fougere's Autoregressive Approach

Fougere [26] tackled the problem of the spectrum analysis of red noise processes (long-memory) using Burg's so-called 'maximum entropy' autoregressive modelling method. Details of Burg's method can be found, for example, in [15, Section 9.5]. When fitting an autoregression of order p the spectral ordinates are estimated from

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

where the estimates of the innovation variance $\tilde{\sigma}_p^2$, and of the parameters $\tilde{\phi}_{j,p}, j = 1, ..., p$ are found using Burg's method. A smooth spectral estimate is produced 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 distributional results, so we cannot easily draw theoretical conclusions about the regression parameter estimates, and simulations are required. The data that Fougere used in his simulations were generated by passing white noise through a FIR filter which produces a power law spectrum over a limited frequency range. Here, we have generated our data using the time-domain fdp-generating method in [12]. For these simulations, the order p of the autoregressive model was fixed at 6, following [26]. The same simulated fdp datasets are used as in Section V, so that we can draw comparisons between all the methods employed.

Table 3 shows the results which may be compared to the multitaper and periodogram results given in Table 2. Burg's algorithm gives better results than the periodogram in terms of estimated mean square error, (bias squared plus variance), except for $d = 0.4, N = 256$, but compared to multitapering its mean square error is much worse for the larger sample size. The bias is almost uniformly worse than when using the periodogram or multitapering. The model order employed, $p = 6$ follows [26], but Fougere admits that "no attempt was made to optimize [the order]" and further that "selection of the correct order [is required] to produce unbiased results, but the order is not known a priori." By way of comparison we chose $K = 5$ for the multitaper method in order to satisfy the theoretical requirement of approximate Gaussianity, a choice which is not data dependent.

B. Likelihood Approach

Although the fdp is a useful parametric model class for investigating the behaviour of various spectral slope estimators, in a practical problem the underlying process is unlikely to be exactly a fdp, and indeed, the data is likely to be somewhat nonGaussian. In this case the nonparametric multitaper spectral estimation approach to estimating β for a spectral decay proportional to $f^{-\beta}$ is still appropriate; it is robust to nonGaussianity through the Central Limit Theorem effect of the Fourier transformations.

Under the idealized case of an exact fdp, then likelihood approaches are appealing. Haslett & Raftery [27] derived an approximate maximum likelihood algorithm for parameter estimation in a full ARFIMA(p, d, q) model. (See [39], [40] for other approximate maximum likelihood techniques). The algorithm may be applied to the more specialised fractionally differenced white noise models, ARFIMA$(0, d, 0)$, that concern us here. A stationary Gaussian stochastic process $\{X_t\}$ with zero mean, will have multivariate Gaussian distribution with conditional mean and variance of X_t, $E\{X_t|X_{t-1}, \ldots, X_1\}$ and $\text{var}\{X_t|X_{t-1}, \ldots, X_1\}$. [27] shows how to accurately approximate these quantities. Along with an initial estimate of d, this enables an approximate maximum likelihood estimate of the innovation variance σ_ϵ^2 to be calculated. The log-likelihood is then formed up with σ_ϵ^2 replaced by its estimate, and the approximate maximum likelihood estimate of d is then found by maximizing this log-likelihood. We used the code made available over the Internet by Haslett and Raftery.

The results for the same fdp data used previously are given in Table 3. The mean square error of the estimates obtained with the approximate maximum likelihood method are comparable with the multitaper estimates for $d = 0.25$, but much larger for $d = 0.4$; they also suffer from a significant negative bias.

Of course, the estimation method matched most closely to the generated fdp data is exact maximum likelihood estimation for an fdp process ([28],[29]). This is computationally quite complex. The *asymptotic* Cramér-Rao lower bound for the variance of an unbiased estimator of d is given in [8] as $6/(N\pi^2)$. For $N = 128$ and 256 this equates to 0.0047 and 0.0024 respectively. We see from Table 2 that for $N = 256$ the multitaper methods have mean square errors close to or less than this. However, Deriche and Tewfik [29] have computed the Cramér-Rao lower bound for some finite sample sizes, and shown that this is somewhat less than the asymptotic bound for larger values of d ($d > 0.25$). In any event, as pointed out earlier, for practical applications the underlying process is unlikely to be exactly a fdp, and is likely to be nonGaussian to some extent, so that exact maximum likelihood based on a fdp model is unlikely to be optimal in practice.

VII. APPLICATION TO GEOPHYSICAL BOREHOLE DATA

A simple, but useful and widely applied model for the earth used in reflection seismology, is that of a stack of horizontal parallel layers of rock, each layer having its own intrinsic acoustic impedance (velocity times density). Between two layers there is an acoustic impedance contrast, corresponding to which is a reflection coefficient, defined as the difference between the acoustic impedances of the two layers divided by their sum. The statistics of the acoustic impedance contrasts (and hence reflection coefficients) with depth, analyzed as a time series, determine how seismic waves propagate through the earth, and knowledge of the nature of the series is a very important factor in understanding the limitations of seismic imaging [5],[30],[31]. The power spectra of the acoustic impedance contrasts, on a logarithmic scale, have been found in numerous studies to be of power-law, i.e., $f^{-\beta}$, type (see e.g., [3],[4]). Since, the value of the spectral decay constant β will determine the nature of the seismic wave propagation, a good estimate of its value is vital.

A logarithmic acoustic impedance series derived from a borehole drilled in a deltaic depositional environment, and comprising more than 1200 values, was analyzed. Multitaper spectral estimation combined with regression on spectral ordinates was used to estimate β. The estimate of β was 0.62 using five tapers, with an estimated standard deviation of 0.02. The estimate was essentially insensitive to the number of tapers, taking the value 0.64 for 8, 10 or 12 tapers. The reduction in the variance of the spectrum ordinates between using one taper and five tapers is illustrated in Figure 7.

VIII. CONCLUSIONS

We have demonstrated that the use of the multitaper spectrum estimation method, combined with regression on the resulting estimated spectrum ordinates provides a sound practical method for estimating a power-law exponent. Note that it is not necessary that the spectrum should be power-law in form over the complete frequency range; the non-parametric methodology is easily adapted to cases where the power-law behaviour holds piecewise over sub-bands of the full frequency range, e.g. [41]. We have shown that the multitaper approach is superior in terms of mean square error to several existing methods. Furthermore, it is robust to nonGaussianity and, with the advent of the sine tapers, is simple to implement.

ACKNOWLEDGEMENT

Emma McCoy was supported through an EPSRC (UK) Grant, and Don Percival by an NSF Phase I SBIR Grant to MathSoft. The authors thank the anonymous referees for numerous useful suggestions leading to an improved exposition.

REFERENCES

[1] W.H. Munk and G.J.F. Macdonald, *The Rotation of the Earth.* Cambridge: Cambridge University Press, 1975.

[2] A.V. Holden, *Models of the Stochastic Activity of Neurones.* Berlin: Springer-Verlag, 1976.

[3] A.T. Walden, and J.W.J. Hosken, "An investigation of the spectral properties of primary reflection coefficients," *Geophys. Prosp.*, vol. 33, pp. 400-435, 1985.

[4] J.P. Todoeschuck and O.G. Jensen, "Joseph geology and seismic deconvolution," *Geophysics*, vol. 53, pp. 1410-1414, 1988.

[5] C. Kerner and P.E. Harris, "Scattering attenuation in sediments modeled by ARMA processes — validation of simple Q models," *Geophysics*, vol. 59, pp. 1813-1826, 1994.

[6] I. McHardy and B. Czerny, "Fractal X-ray time variability and spectral invariance of the Seyfert galaxy NGC 5506," *Nature*, vol. 325, pp. 696-698, 1987.

[7] J. Geweke and S. Porter-Hudak, "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, vol. 4, pp. 221-238, 1983.

[8] R.L. Kashyap and K-B. Eom, "Estimation in long memory time series models," *Journal of Time Series Analysis*, vol. 9, pp. 35-41, 1988.

[9] J.R.M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165-176, 1981.

[10] D.R. Brillinger, *Time Series: Data Analysis and Theory, Expanded Edition.* New York: McGraw-Hill, 1981.

[11] M.B. Priestley, *Spectral Analysis and Time Series.* New York: Academic Press, 1981.

[12] J.R.M. Hosking, "Modeling persistence in hydrological time series using fractional differencing," *Water Res. Res.*, vol. 20, pp. 1898-1908, 1984.

[13] M.C. Viano, Cl. Deniau, and G. Oppenheim, "Long-range dependence and mixing for discrete time fractional processes," *Journal of Time Series Analysis*, vol. 16, pp. 323-338, 1995.

[14] D.J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055-1096, 1982.

[15] D.B. Percival and A.T. Walden, *Spectral Analysis for Physical Applications.* Cambridge: Cambridge University Press, 1993.

[16] A.T. Walden, E.J. McCoy and D.B. Percival, "The variance of multitaper spectrum estimates for real Gaussian processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 479-482, 1994.

[17] D.J. Thomson, "Quadratic-inverse spectrum estimates: applications to palaeoclimatology," *Phil. Trans. Roy. Soc. London, Series A*, vol. 332, pp. 539-597, 1990.

[18] K.S. Riedel and A. Sidorenko, "Mininum biased multitaper spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 188-195, 1995.

[19] B. Bell, D.B. Percival and A.T. Walden, "Calculating Thomson's spectral multitapers by inverse iteration," *J. Comp. Graph. Stat.*, vol. 2, pp. 119-130, 1993.

[20] A.T. Walden, E.J. McCoy and D.B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, pp. 201-214, 1995.

[21] M.S. Bartlett and D.G. Kendall, "The statistical analysis of variance- heterogeneity and the logarithmic transformation," *Supplement to the Journal of the Royal Statistical Society*, vol. 8, pp. 128-138, 1946.

[22] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods, 2nd Edition.* New York: Springer-Verlag, 1991.

[23] V.A. Reisen, "Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram," *Journal of Time Series Analysis*, vol. 15, pp. 335-350, 1994.

[24] G. Chen, A. Bovas, and S. Peiris, "Lag window estimation of the degree of differencing in fractionally integrated time series models," *Journal of Time Series Analysis*, vol. 15, pp. 473-487, 1994.

[25] C.W.J. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series.* Princeton: Princeton University Press, 1964.

[26] F. Fougere, "On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods," *Journal of Geophysical Research*, vol. 90, pp. 4355-4366, 1985.

[27] J. Haslett and A.E. Raftery, "Space-time modelling with long-memory dependence: assessing Ireland's wind power resource," *Applied Statistics,* vol. 38, pp. 1-50, 1989.

[28] J.B. Carlin, A.P. Dempster, and A.B. Jonas, "On models and methods for Bayesian time series analysis," *Journal of Econometrics*, vol. 30, pp. 67-90, 1985.

[29] M. Deriche and A.H. Tewfik, "Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process," *IEEE Trans. Signal Processing*, vol. 41, pp. 2977-2989, 1993.

[30] J.W.J. Hosken, S.K. Mayo, and A.T. Walden, "The influence of reflection coefficient statistics on the seismic method: scattering attenuation and transmission wavelets," *Proc. Roy. Soc. London, Series A.*, vol. 439, pp. 1-23, 1992.

[31] J.W.J. Hosken, S.K. Mayo, and A.T. Walden, "The influence of reflection coefficient statistics on the seismic method: scattering and the minimum detectable reflection coefficient," *Proc. Roy. Soc. London, Series A.*, vol. 440, pp. 221-240, 1993.

[32] Y. Yajima, "A central limit theorem of Fourier transforms of strongly dependent stationary processes," *Journal of Time Series Analysis*, vol. 10, pp. 375-383, 1989.

[33] C.M. Hurvich and K.I. Beltrao, "Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series," *Journal of Time Series Analysis*, vol. 14, pp. 455-472, 1993.

[34] P.M. Robinson, "Log-periodogram regression on time series with long range dependence," *The Annals of Statistics*, vol. 23, pp. 1048-1072, 1995.

[35] A.T. Walden, D.B. Percival and E.J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," Submitted to *IEEE Transactions on Signal Processing*, 1997.

[36] K.S. Riedel and A. Sidorenko, "Adaptive smoothing of the log-spectrum with multiple tapering," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1794-1800, 1996.

[37] C. Agiakloglou, P. Newbold and M. Wohar, "Bias in an estimator of the fractional difference parameter," *Journal of Time Series Analysis*, vol. 14, pp. 235-246, 1993.

[38] C.M. Hurvich and K.I. Beltrao, "Automatic semiparametric estimation of the memory parameter of a long-memory time series," *Journal of Time Series Analysis*, vol. 15, pp. 285-302, 1994.

[39] J. Beran, "Fitting long-memory models by generalized linear regression," *Biometrika*, vol. 80, pp. 817-822, 1993.

[40] J. Beran, "Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models," *J. Royal Statistical Society, B*, vol. 57, pp. 659-672, 1995.

[41] D.B. Percival, "Characterization of frequency stability: frequency-domain estimation of stability measures," *Proc. IEEE*, vol. 79, pp. 961-972.

Type	$N = 128$	$N = 256$
Periodogram	0.1006	0.0677
Ordinary LS, 4 sine tapers	0.0852	0.0593
Ordinary LS, 5 sine tapers	0.0810	0.0567
Ordinary LS, 6 sine tapers	0.0779	0.0549
Weighted LS, 4 sine tapers	0.0694	0.0514
Weighted LS, 5 sine tapers	0.0626	0.0473
Weighted LS, 6 sine tapers	0.0574	0.0439

Table 1: Comparison of theoretical standard deviations of least squares estimators of d using multitapering with 4, 5 or 6 sine tapers, and the periodogram.

d		periodogram		dpss tapers		sine tapers (ord. ls)		sine tapers (wgt. ls)	
		$N = 128$	$N = 256$	$N = 128$	$N = 256$	$N = 128$	$N = 256$	$N = 128$	$N = 256$
0.25	bias	0.0136	0.0048	0.0217	0.0055	0.0238	0.0049	0.0217	0.0021
	std. error	0.1030	0.0681	0.0843	0.0515	0.0790	0.0496	0.0775	0.0472
0.4	bias	-0.0121	-0.0031	0.0181	0.0092	0.0113	0.0063	0.0082	0.0037
	std. error	0.1179	0.0668	0.0848	0.0418	0.0686	0.0436	0.0609	0.0417

Table 2: Bias and standard errors of 500 estimates of d obtained from regression on periodogram and 5-taper multitaper spectrum ordinates.

d		Burg's method		approx. max. lik.		sine tapers (ord. ls)	
		$N = 128$	$N = 256$	$N = 128$	$N = 256$	$N = 128$	$N = 256$
0.25	bias	0.0230	0.0201	-0.0451	-0.0215	0.0238	0.0049
	std. error	0.0804	0.0574	0.0752	0.0523	0.0790	0.0496
0.4	bias	0.0462	0.0510	-0.0567	-0.0275	0.0113	0.0063
	std. error	0.0810	0.0624	0.0706	0.0484	0.0686	0.0436

Table 3: Bias and standard errors of 500 estimates of d obtained using Fougere's method based on Burg's algorithm, Haslett and Raftery's approximate maximum likelihood, and regression on 5-taper multitaper spectrum ordinates.

LIST OF FIGURES

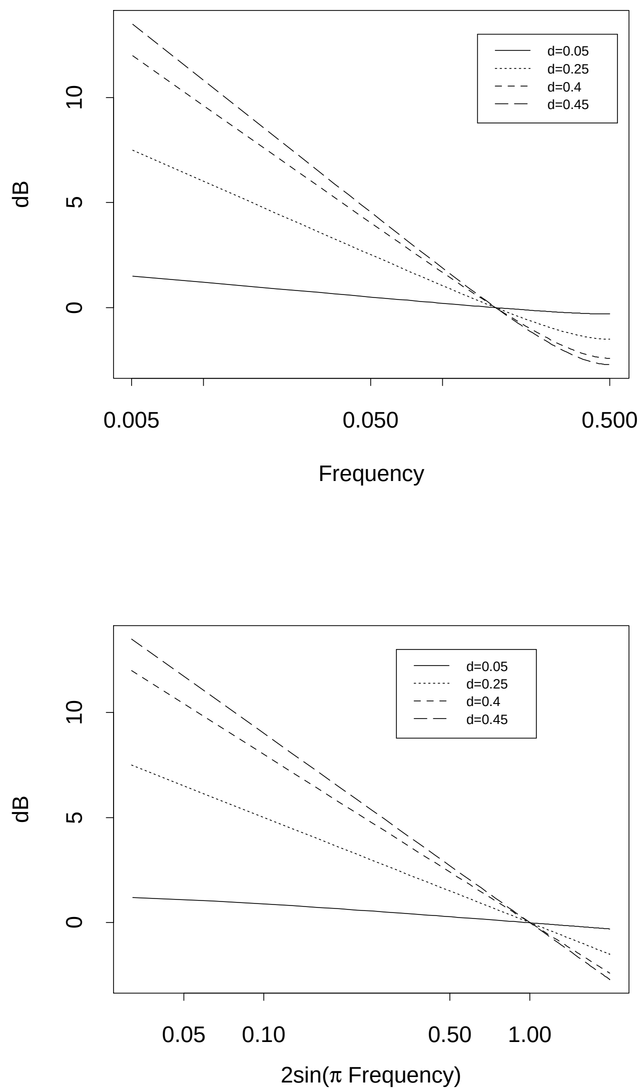


Figure 1a (upper), 1b (lower).

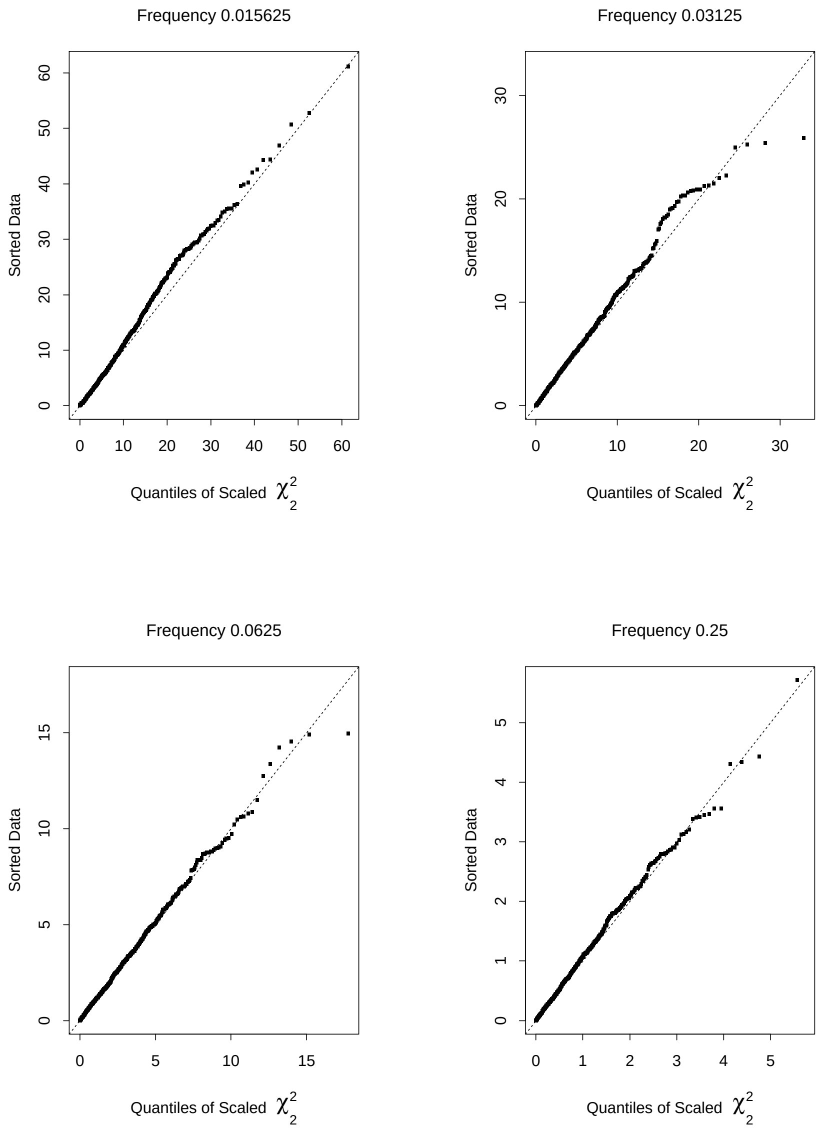

Frequency 0.015625
Frequency 0.03125
Frequency 0.0625
Frequency 0.25
Sorted Data
Quantiles of Scaled χ^2_2
0
10
20
30
40
50
60
0
5
10
15
1
2
3
4
5

Figure 2.

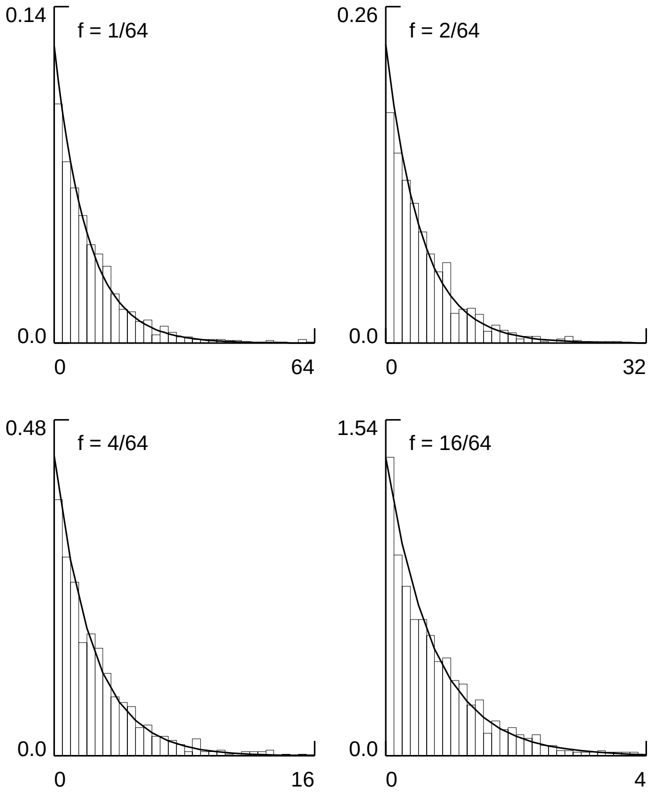


Figure 3.

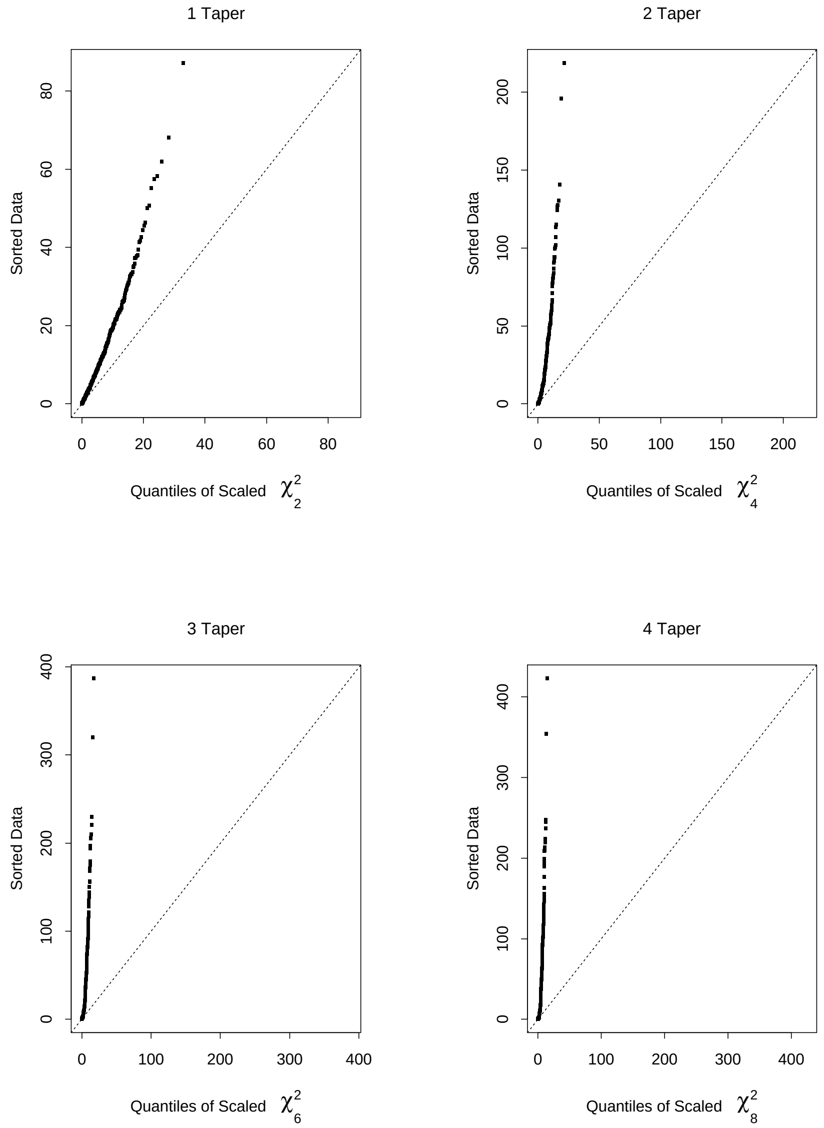

1 Taper
Sorted Data
Quantiles of Scaled χ^2_2
2 Taper
Sorted Data
Quantiles of Scaled χ^2_4
3 Taper
Sorted Data
Quantiles of Scaled χ^2_6
4 Taper
Sorted Data
Quantiles of Scaled χ^2_8

Figure 4a.

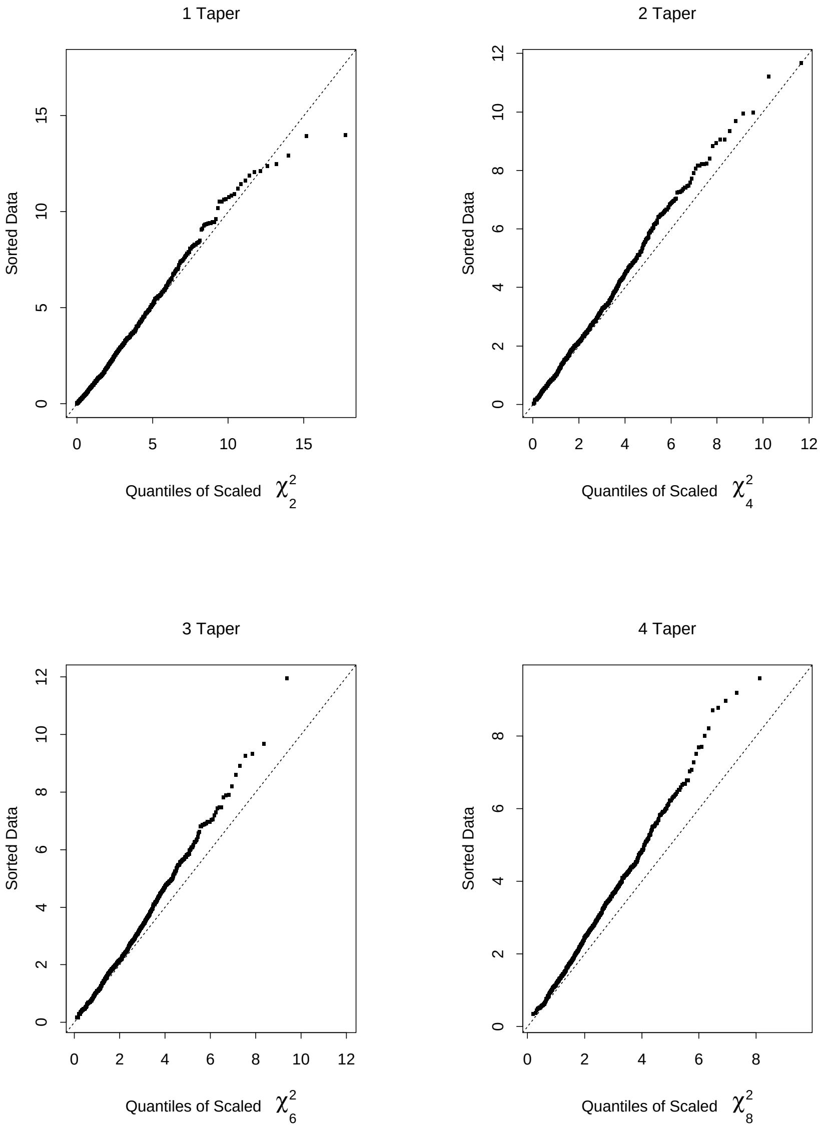

1 Taper
Sorted Data
Quantiles of Scaled χ^2_2
2 Taper
Sorted Data
Quantiles of Scaled χ^2_4
3 Taper
Sorted Data
Quantiles of Scaled χ^2_6
4 Taper
Sorted Data
Quantiles of Scaled χ^2_8

Figure 4b.

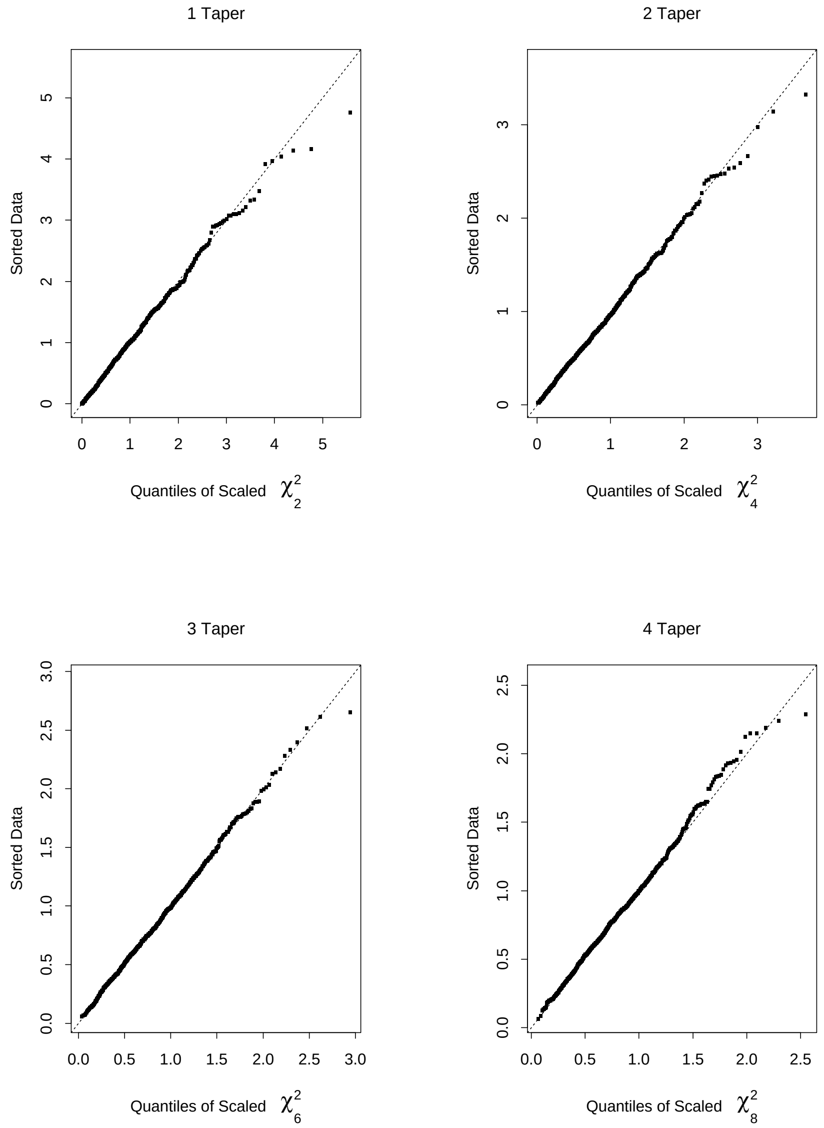

1 Taper
Sorted Data
Quantiles of Scaled χ^2_2
2 Taper
Sorted Data
Quantiles of Scaled χ^2_4
3 Taper
Sorted Data
Quantiles of Scaled χ^2_6
4 Taper
Sorted Data
Quantiles of Scaled χ^2_8

Figure 4c.

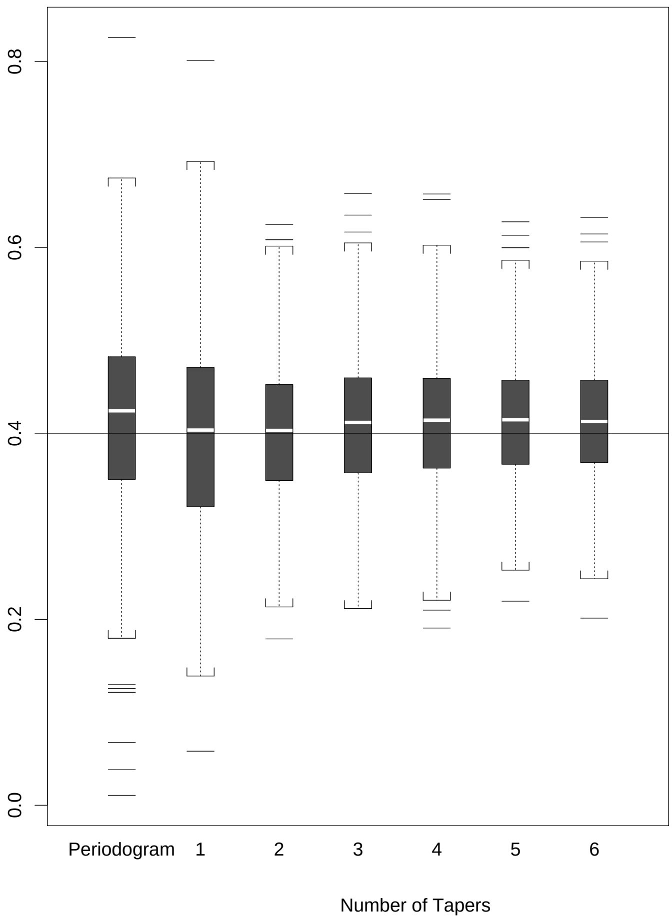
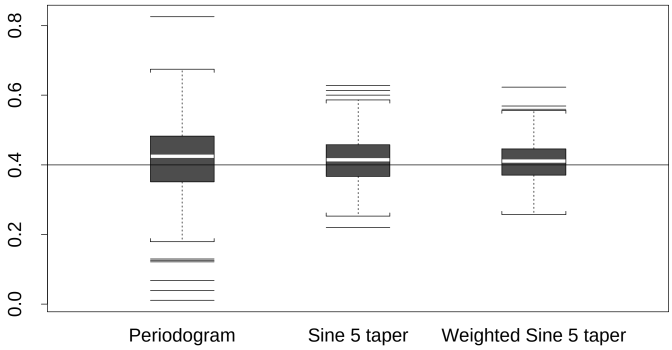
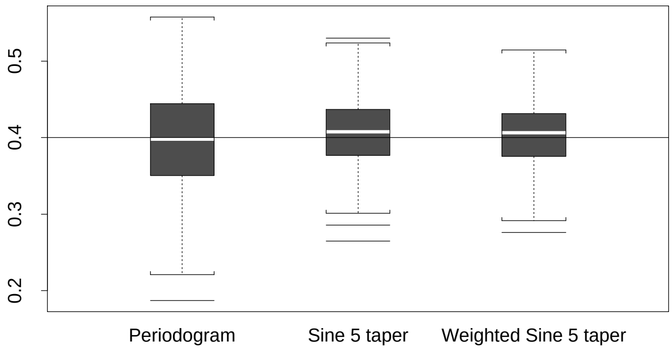


Figure 5.



N=128



N=256

Figure 6

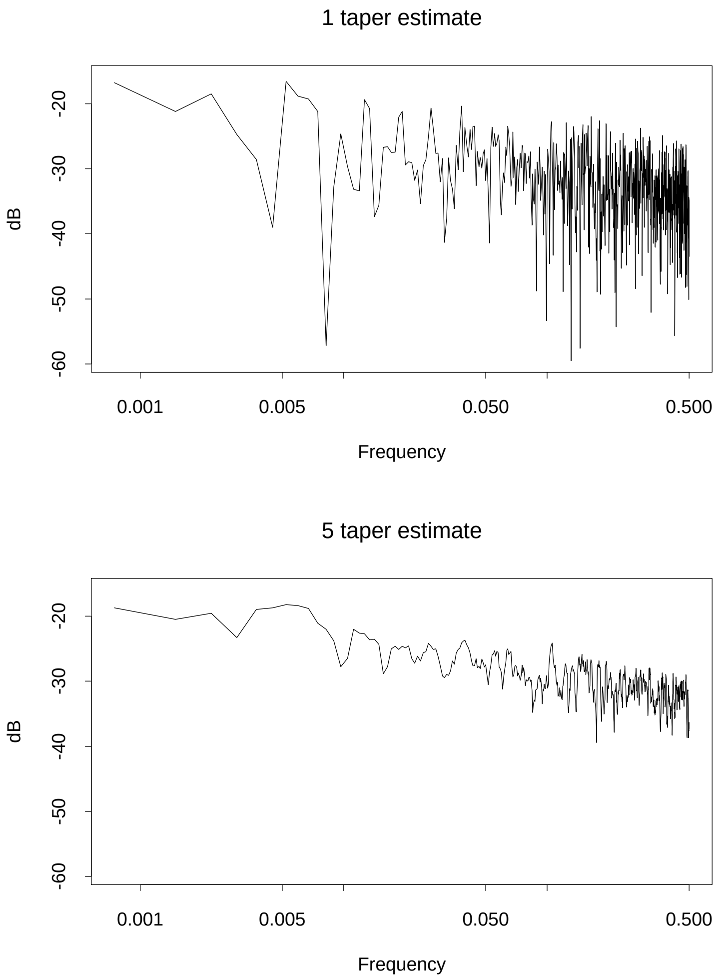

1 taper estimate
dB
-20
-30
-40
-50
-60
0.001
0.005
0.050
0.500
Frequency
5 taper estimate
dB
-20
-30
-40
-50
-60
0.001
0.005
0.050
0.500
Frequency

Figure 7.