BIAS-CORRECTED NONPARAMETRIC SPECTRAL ESTIMATION II

by

Dimitris N. Politis and Joseph P. Romano
Purdue University Stanford University

Technical Report #94-5

Department of Statistics
Purdue University

March 1994

Bias-corrected nonparametric spectral estimation II

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

The theory of nonparametric spectral density estimation based on an observed stretch $X_1, \ldots, X_N$ from a stationary time series has been studied extensively in recent years. However, the most popular spectral estimators, such as the ones proposed by Bartlett, Daniell, Parzen, Priestley, and Tukey, are plagued by the problem of bias, which effectively prohibits $\sqrt{N}$-convergence of the estimator. This is true *even* in the case the data are known to be m-dependent in which case $\sqrt{N}$-consistent estimation is possible by a simple plug-in method.

In this report, an intuitive method for the reduction in the bias of a nonparametric spectral estimator is presented. In fact, applying the proposed methodology to Bartlett's estimator results in bias-corrected estimators that are related to kernel estimators with lag-windows of trapezoidal shape. The asymptotic performance (bias, variance, rate of convergence) of the proposed estimators is investigated; in particular, it is found that the trapezoidal lag-window spectral estimator is $\sqrt{N}$-consistent in the case of MA processes, and $\sqrt{N/\log N}$-consistent in the case of ARMA processes. The finite-sample performance of the trapezoidal lag-window estimator is also assessed by means of a numerical simulation.

Keywords. Bartlett's estimator, bias reduction, mean squared error, lag-windows, nonparametric spectral estimation, variance estimation.

1. Introduction

Suppose $X_1, \ldots, X_N$ are observations from a stationary time series $\{X_t, t \in Z\}$ with mean zero, i.e., $EX_t = 0$. Assume that the spectral density function $f(w)$ exists, and is defined by $f(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} R(s)e^{-jsw}$, where $R(s) = EX_tX_{t+|s|}$ is the autocovariance; note that the symbol j denotes the imaginary unit $\sqrt{-1}$. Attention will focus on nonparametric estimators of the spectral density function $f(w)$, and on designing such estimators with small bias.

One of the first (and most intuitive) proposals for consistent estimation of $f(w)$ at some point $w \in [-\pi, \pi]$ was given by Bartlett (1946) and is defined by

$$\tilde{f}(w) = \frac{1}{2\pi}\sum_{s=-M}^{M}(1 - \frac{|s|}{M})\hat{R}(s)e^{-jws} \tag{1}$$

where $\hat{R}(s) = \frac{1}{N}\sum_{k=1}^{N-|s|} X_kX_{k+|s|}$ is the usual sample autocovariance. The Bartlett estimator $\tilde{f}$ can alternatively be computed as an average of short periodograms (cf. Priestley (1981)), in which case its computation is most efficient (cf. Gardner (1988)). Under regularity conditions (cf. Priestley(1981), Zhurbenko (1986)), $\tilde{f}(w)$ is a consistent and asymptotically normal estimator of the spectral density function $f(w)$. The regularity conditions are, roughly speaking, moment conditions, weak dependence conditions (that sometimes are expressed by conditions on the smoothness of f), and conditions on the design parameter M.

In particular, the asymptotic variance of $\tilde{f}(w)$ is given by

$$Var(\tilde{f}(w)) \sim \frac{2}{3}\frac{M}{N}f^2(w)(1 + \eta(w)), \tag{2}$$

under some conditions[1], where $\eta(w) = 0$ if $w \neq 0(mod\pi)$ and $\eta(w) = 1$ if $w = 0, \pm\pi$; following the usual convention, the notation $A_N \sim B_N$ is a short-hand for $A_N/B_N \to 1$, as $N \to \infty$. In addition, the large sample distribution of $\sqrt{N/M}(\tilde{f}(w) - f(w))$ is the normal $N(0, \frac{2}{3}f^2(w)(1 + \eta(w)))$, provided $M \sim aN^{\beta}$, for some $a > 0$ and $\beta > 1/3$, as $N \to \infty$.

[1]There is a variety of sufficient conditions guaranteeing that equation (2) holds; for example (cf. Priestley (1981)) a sufficient condition is that $\{X_t\}$ is a linear process given by $X_t = \sum_{i=-\infty}^{\infty}\theta_iZ_{t-i}$, where the Z_t's are i.i.d. with $EZ_t = 0$, and $EZ_t^4 < \infty$, and the θ_i's are constants satisfying $\sum_{i=-\infty}^{\infty}|i|^{1/2}|\theta_i| < \infty$. For different sufficient conditions that are based on summability of cumulants and do not require that $\{X_t\}$ is a linear process see Brillinger (1975, p.26 and p. 144), or Rosenblatt (1985, p.134).

It can also be calculated that the bias of $\tilde{f}(w)$ is of approximate order $c_{w,f}/M$, for some $c_{w,f} \neq 0$ depending on w and on f, under the regularity condition $\sum_{s=-\infty}^{\infty} |s||R(s)| < \infty$; since the order of magnitude of the bias is rather large, subsequent research efforts were expended to obtain spectral estimators with smaller bias. These efforts pointed to the direction of using a different kernel that is smoother near the origin than Bartlett's triangular one (cf. Grenander and Rosenblatt(1957), Blackman and Tukey(1959), Parzen(1957a,b), Priestley(1981)). By appropriately choosing the kernel, and assuming that f has a continuous second derivative, one can have an estimator that is nonnegative, and possesses a bias of approximate order $O(1/M^2)$, which is a significant improvement for large samples.

In this paper, a different perspective on the problem of bias reduction will be presented, and a new way to look at such smoothing problems will be discussed. This will result in a specific proposal for bias-corrected estimates that are related to kernel type estimators with lag-windows of trapezoidal shape. These bias-corrected estimates will be shown to possess very small bias of approximate order $O(1/M^r)$, while having variance of order $O(M/N)$ as usual; here r can be intuitively interpreted as the number of derivatives that f has.

Attention will focus on Bartlett's estimator because it is both most intuitive and simple to calculate, and is the most needy for a bias correction; in addition, Bartlett's estimator (evaluated at point $w = 0$) comes up rather naturally as a variance estimate in the newly developed areas of resampling and subsampling dependent data (cf. Künsch (1989), Liu and Singh (1992), Politis and Romano (1992a,c)), and in the steady state simulation literature (cf. Meketon and Schmeiser (1984), Welch (1987), Song and Schmeiser (1988, 1992)). With the appropriate modifications, the same intuitive procedure can be carried out for bias reduction of other spectral estimators. However, as will be apparent later on, bias-correcting Bartlett's estimator has the unique effect of taking the bias all the way down to order $O(1/M^r)$; applying the bias-correction procedure to a different estimator with bias of order $O(1/M^q)$, with $q \geq 2$, would only bring the bias of the corrected estimator to order $O(1/M^{q+2})$, which is bigger than $O(1/M^r)$ if $r > q+2$.

The remaining of the paper is organized as follows. In Sections 2.1 and 2.2, the main intuitive proposal for bias reduction is presented, and its good asymptotic properties are established. Section 2.3 is concerned with the positivity of the proposed estimator, and Sections 2.4 and 2.5 establish the rates of convergence in the case of ARMA and MA models respectively. Section 2.6 addresses the most important problem of choosing the bandwidth of the spectral estimator in practice, and Section 2.7 discusses the use of the unbiased autocovariances in forming the spectral estimator. Finally, Section 3 presents some numerical finite-sample results, and Section 4 contains some further comments and conclusions. All technical proofs are placed in the Appendix.

2. Bias reduction for Bartlett's spectral estimator

2.1. The notion of bias-correction. From the data $X_1, \ldots, X_N$ construct the Bartlett spectral estimator $\tilde{f}(w)$ at point w as given in equation (1) with *some* choice of M; the optimal choice of M will be discussed later. Also construct an over-smoothed Bartlett spectral estimator $\bar{f}(w)$ using a different block size $\bar{m} < M$; $\bar{f}(w)$ can be thought of as a crude estimate of $f(w)$.

Looking at $f(w), E\tilde{f}(w), E\bar{f}(w)$ as functions of $w \in [-\pi, \pi]$ it is obvious (cf. Priestley (1981)) that $E\tilde{f}(w)$ is a smoothed version of $f(w)$, and, in turn, $E\bar{f}(w)$ is a smoothed version of $E\tilde{f}(w)$. This observation leads to the heuristic approximation

$$Bias(\tilde{f}(w)) \equiv E\tilde{f}(w) - f(w) \simeq E\bar{f}(w) - E\tilde{f}(w). \tag{3}$$

If the approximation (3) were somehow correct, then we could estimate $E\tilde{f}(w)$ by $\tilde{f}(w)$ and $E\bar{f}(w)$ by $\bar{f}(w)$, and we could therefore estimate the bias of $\tilde{f}(w)$ by $\widehat{Bias}(\tilde{f}(w)) = \bar{f}(w) - \tilde{f}(w)$. As a consequence, we could form a 'bias-corrected' Bartlett's estimator by the formula

$$\hat{f}(w) \equiv \tilde{f}(w) - \widehat{Bias}(\tilde{f}(w)) = 2\tilde{f}(w) - \bar{f}(w). \tag{4}$$

Notably this proposed bias-correction is in the spirit of Quenouille's (1949) original suggestion of bias reduction for time series statistics. In Quenouille's scheme, a crude estimate of $f(w)$ would be obtained by first splitting the series $X_1, \ldots, X_N$ into, say, two subseries of length $N/2$, secondly constructing Bartlett estimates from each subseries (obviously using a different block size $\bar{m}$, smaller than the original M), and finally averaging the two estimates arising from the two subseries. The bias-corrected estimate would then be calculated by subtracting this crude estimate from twice the original Bartlett estimate $\tilde{f}(w)$. In our proposal, the crude estimate $\bar{f}(w)$ is obtained in a slightly more general fashion, employing the whole time series anew.

Of course, it is most optimistic to expect that the approximation (3) would hold. However, the following heuristic argument indicates that (3) is at least qualitatively true, in which case $\hat{f}(w)$ might still have reduced bias as compared to $\tilde{f}(w)$. Suppose that the true spectral density $f(w)$ has a peak around w_0; for example, see Figure 4 (a) that is related to the simulation experiment of Section 3, and where $w_0 \simeq 0.88$. In this case, the bias of $\tilde{f}(w)$ is intuitively due to either 'smoothing out' the peak (for w close to w_0), or to 'leakage' from the peak (for w away from w_0). It is then easy to see that for w close to w_0, both quantities, $\tilde{f}(w) - f(w)$ and $\bar{f}(w) - \tilde{f}(w)$ are negative; similarly, for w away from w_0, both $\tilde{f}(w) - f(w)$, as well as $\bar{f}(w) - \tilde{f}(w)$ are positive.

In other words, the proposed simple bias estimate $\widehat{Bias}(\tilde{f}(w)) = \bar{f}(w) - \tilde{f}(w)$ will capture the sign of $Bias(\tilde{f}(w))$, although not necessarily its absolute value. Hence, a reasonable next step is to approximate $Bias(\tilde{f}(w))$ by $h\widehat{Bias}(\tilde{f}(w))$, where h is a positive constant to be specified later. So, by defining $\widehat{\widehat{Bias}}(\tilde{f}(w)) = h\widehat{Bias}(\tilde{f}(w)) = h(\bar{f}(w) - \tilde{f}(w))$, we are led to a general bias-corrected estimator given by

$$\hat{f}_h(w) \equiv \tilde{f}(w) - \widehat{\widehat{Bias}}(\tilde{f}(w)) = \tilde{f}(w) - h\widehat{Bias}(\tilde{f}(w)) = (h+1)\tilde{f}(w) - h\bar{f}(w). \tag{5}$$

It is interesting to note that empirical results of Song and Schmeiser (1988) concerning estimation of $f(w)$ at the point $w = 0$, pointed to a formula analogous to (5) as the linear combination of spectral estimators with minimum mean squared error.

Since, as mentioned in the Introduction, $E\tilde{f}(w) = f(w) + c_{w,f}/M + o(1/M)$, and $E\bar{f}(w) = f(w) + c_{w,f}/\bar{m} + o(1/\bar{m})$, it is intuitively clear that letting $h = \bar{m}/(M - \bar{m})$ yields $E\hat{f}(w) = f(w) + o(1/M)$, so that the estimator bias is actually reduced from $O(1/M)$ to $o(1/M)$; however, the bias reduction is actually quite more dramatic, as Theorem 1 will demonstrate later on. It is to be noted though, that by putting $h = \bar{m}/(M - \bar{m})$, the proposed bias-corrected estimator is actually a nonparametric spectral estimator of the lag-window type, i.e.,

$$\hat{f}_h(w) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \lambda(s)\hat{R}(s)e^{-jsw}, \tag{6}$$

where the lag-window $\lambda(s)$ has the general trapezoidal shape given below, i.e.,

$$\lambda(s) = \begin{cases} 1 & \text{for } |s| \le \bar{m} \\ 1 - \frac{|s|-\bar{m}}{M-\bar{m}} & \text{for } \bar{m} < |s| \le M \\ 0 & \text{for } |s| > M. \end{cases}$$

A more precise notation for the above trapezoidal lag-window might be $\lambda_{h,M}(s)$, where $h = \bar{m}/(M - \bar{m})$, but we will generally drop the subscripts h, M since it does not lead to confusion.

2.2. Performance of the general trapezoidal lag-window. As elaborated upon in Section 2.1, the estimator $\hat{f}_h(w)$ can be computed as a linear combination of the two Bartlett estimators $\tilde{f}(w)$ and $\bar{f}(w)$, by the formula

$$\hat{f}_h(w) = (h+1)\tilde{f}(w) - h\bar{f}(w) = \frac{M}{M-\bar{m}}\tilde{f}(w) - \frac{\bar{m}}{M-\bar{m}}\bar{f}(w), \tag{7}$$

with $h = \bar{m}/(M-\bar{m})$. The assumption that h is a constant is equivalent to assuming that $\bar{m}$ is proportional to M, i.e., $\bar{m} \sim cM$, for some constant $c \in (0,1)$. Nevertheless, varying $h \in (0,\infty)$ a whole family of bias-corrected estimators is obtained; if $h = 1$, then the estimator $\hat{f}_h(w)$ is identical to $\hat{f}(w)$ defined in equation (4). Observe that at the extreme point where $h = 0$, i.e., $\bar{m} = 0$, $\hat{f}_h(w)$ reduces to a Bartlett estimator, and at the extreme point where $h = \infty$, i.e., $\bar{m} = M$, $\hat{f}_h(w)$ becomes a 'truncated periodogram' $\frac{1}{2\pi}\sum_{s=-M}^{M}\hat{R}(s)e^{-jsw}$.

It should be noted (cf., for example, Brockwell and Davis (1991)) that the estimator $\hat{f}_h(w)$ can *also* be written as $\hat{f}_h(w) = \int_{-\pi}^{\pi}\Lambda_h(w')I_N(w+w')dw'$, where $I_N(w) = \frac{1}{2\pi}\sum_{s=-N+1}^{N-1}\hat{R}(s)e^{-jsw}$ is the periodogram, and

$$\Lambda_h(w) \equiv \frac{1}{2\pi}\sum_{s=-M}^{M}\lambda(s)e^{-jsw} \tag{8}$$

is the so-called spectral window corresponding to the lag-window $\lambda(s)$. By the previous discussion, and since

$$\Lambda_0(w) = \frac{1}{2\pi M}\left(\frac{\sin(Mw/2)}{\sin(w/2)}\right)^2$$

is the well-known Fejér spectral kernel corresponding to the Bartlett estimator $\tilde{f}(w)$, we have an explicit formula for $\Lambda_h(w)$, that is,

$$\begin{aligned}\Lambda_h(w) &= \frac{h+1}{2\pi M}\left(\frac{\sin(Mw/2)}{\sin(w/2)}\right)^2 - \frac{h}{2\pi\bar{m}}\left(\frac{\sin(\bar{m}w/2)}{\sin(w/2)}\right)^2 \\ &= \frac{\sin^2(Mw/2) - \sin^2(\bar{m}w/2)}{2\pi(M-\bar{m})\sin^2(w/2)}\end{aligned} \tag{9}$$

Since the extreme case $h = \infty$ corresponds to the truncated periodogram, it follows that $\Lambda_\infty(w) = \sin(\frac{(2M+1)w}{2})/2\pi\sin(w/2)$ is the familiar Dirichlet kernel. Note that, in view of representation (9), the spectral kernel $\Lambda_h(w)$, for $0 \le h < \infty$ possesses many interesting properties that are summarized in the following lemma.

Lemma 1 *If $h \in [0, \infty)$, then*

(a) $\Lambda_h(w)$ is an even function of w, with period 2π;

(b) $\int_{-\pi}^{\pi} \Lambda_h(w) = 1$;

(c) for any $\epsilon > 0$, $\int_{-\epsilon}^{\epsilon} \Lambda_h(w) \to 1$, as $M \to \infty$.

(d) If $h \in (0, \infty)$, and k is any even positive integer, then $\int_{-\pi}^{\pi} w^k \Lambda_0(w)dw \sim b_k/M$, while $\int_{-\pi}^{\pi} w^k \Lambda_h(w)dw = O(1/M^2)$, as $M \to \infty$, where b_k is a nonzero constant depending on k only.

Property (d) above points to another way of seing that that $\hat{f}_h(w)$ has smaller bias than $\hat{f}_0(w) = \tilde{f}(w)$. For example, if $\sum_{s=-\infty}^{\infty} |s|^3 |R(s)| < \infty$, then by corollary 5.8.2 of Brillinger (1975),

$$E\hat{f}_h(w) = f(w) + \frac{1}{2M^2}\left(\int_{-\pi}^{\pi} t^2 \Lambda_h(t)dt\right) f^{(2)}(w) + O(1/M^3) + O(M/N),$$

where $f^{(2)}(w)$ is the 2nd derivative of $f(w)$; it follows that the $Bias(\hat{f}_h(w)) = O(1/M^3)$, if $M \to \infty$, as $N \to \infty$, but with $M^4/N \to 0$.

Much more is true however, and the following theorem quantifies the bias-variance performance of $\hat{f}_h$ as an estimator of f. If f is smooth enough, the estimator $\hat{f}_h$ is shown to have smaller (by orders of magnitude) bias than the Bartlett estimator $\tilde{f}$, while its variance remains of the same order of magnitude.

Theorem 1 *Assume that $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$, for some positive integer r. Also assume that $\bar{m} \sim cM$, for some constant $c \in (0,1)$, and that $M \to \infty$, as $N \to \infty$, but with $M^r/N \to 0$. Then*

$$\sup_{w \in [-\pi,\pi]} |Bias(\hat{f}_h(w))| = o(1/M^r). \tag{10}$$

Suppose in addition that the time series $\{X_t\}$ is such that equation (2) holds. Then we also have

$$Var(\hat{f}_h(w)) \sim \frac{3h+1}{h+1}\left(\frac{2M}{3N}\right) f^2(w)(1+\eta(w)) \equiv (2c+1)\left(\frac{2M}{3N}\right) f^2(w)(1+\eta(w)). \tag{11}$$

The interpretation of Theorem 1 is that, for the case $r > 1$, the bias of $\hat{f}_h(w)$ is smaller than the bias of the Bartlett estimator $\tilde{f}(w)$ by orders of magnitude. In particular, if f is smooth enough, (i.e., if r is large), the bias of $\hat{f}_h(w)$ can be considered negligible even for moderately large M. From the theorem's proof it is obvious that the small bias of $\hat{f}_h$ is a consequence of the 'flatness' of the trapezoidal lag-window $\lambda(s)$ for $|s| < \bar{m}$. Indeed, one can conceive of other lag-windows with a 'flat-top' that have favorable bias performance, but we will focus on the trapezoidal one because it arises naturally in the process of bias-correcting the Bartlett window, and it is the simplest to describe and use; the reader is referred to Devroye (1987) for a discussion on 'flat-top' kernels in the analogous context of probability density estimation.

It should be noted that the assumption $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$ implies that f has r continuous derivatives; conversely, if f has r derivatives and the rth derivative satisfies a uniform Lipschitz condition of degree $\alpha > 1/2$, then $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$ follows (cf. Katznelson (1968)). An important class of time series possessing very smooth spectral densities is the family of ARMA models; see the discussion after Theorem 15 in Section 2.4.

Under the assumptions of Theorem 1, the estimator $\hat{f}_h(w)$ is consistent for $f(w)$, since its mean squared error $MSE(\hat{f}_h(w)) \equiv E(\hat{f}_h(w) - f(w))^2 \to 0$, as $N \to \infty$. Returning now to the deferred question of proper choice of M, it follows that by letting $M \sim aN^{1/(2r+1)}$, for some constant $a > 0$, the $MSE(\hat{f}_h(w))$ becomes of order $O(N^{-2r/(2r+1)})$. One can further try to choose the constant a to also minimize the proportionality constant in $MSE(\hat{f}_h(w)) = O(N^{-2r/(2r+1)})$, although this is quite more difficult and will not be pursued here.

It is interesting to observe that the bias of $\hat{f}_h(w)$ will be of asymptotic order $o(1/M^r)$ *regardless* of the choice of h (or, equivalently, of the choice of $\bar{m}$). This goes to show that the choice of h is not as important as the choice of M, and thus the '$2f - f$' rule of equation (4) might be preferred in practice to the more general formula (7) in view of its simplicity. What is, of course, of great importance concerning the design of a spectral estimator, is the choice of M in practical applications; this problem will be taken up again in Section 2.6.

2.3. Taking the positive part. A question that has been overlooked until now is whether the estimator $\hat{f}_h(w)$ is nonnegative or not. Since the spectral density is nonnegative, it is quite important that its estimators be nonnegative as well. Following Parzen (1957a), we define the 'characteristic exponent' of a lag-window $k(s)$ to be the largest positive integer k_0 such that $\lim_{s\to 0} \frac{1-k(s)}{|s|^{k_0}}$ exists, is finite, and is non-zero. If $\lim_{s\to 0} \frac{1-k(s)}{|s|^{k_0}}$ exists for any positive integer k_0, the characteristic exponent is said to be ∞. It is apparent that the characteristic exponent quantifies the smoothness of the lag-window near the origin.

Classifying all lag-windows according to their characteristic exponent yields the following insights: (a) lag-windows with characteristic exponent equal to 1 (e.g., the Bartlett triangular lag-window) lead to heavily biased spectral estimates; and (b) lag-windows with characteristic exponent greater than 2 correspond to spectral kernels that are not everywhere nonnegative, and hence may lead to negative spectral density estimates (cf. Priestley (1981, p. 568)). In view of this, the focus of researchers in the spectral estimation literature has been focused to those lag-windows with characteristic exponent equal to 2 that correspond to nonnegative spectral kernels (cf. Priestley (1981, p. 463) for a list of examples).

However, it is easy to see that the characteristic exponent of the trapezoidal lag-window $\lambda(s)$ is ∞, and hence $\hat{f}_h(w)$ is *not* necessarily nonnegative. To intuitively see this, consider $\hat{f}_h(w)$ with $h = 1$, i.e., $\hat{f}_1(w) = 2\tilde{f}(w) - \bar{f}(w)$, with $\bar{m} = M/2$. It is apparent that the spectral (Fejér) kernel of $\tilde{f}(w)$, i.e., $\Lambda_0(w)$, has twice as many zeros as the spectral kernel of $\bar{f}(w)$; consequently, at the location of a zero of $\Lambda_0(w)$ that is not a zero of the spectral kernel of $\bar{f}(w)$, the spectral kernel of $\hat{f}_1(w)$, i.e., $\Lambda_1(w)$, goes negative. In Figure 1, the spectral kernel $\Lambda_0(w)$ is plotted (for $M = 40$), while in Figure 2, the spectral kernel $\Lambda_1(w)$ is shown. For further comparison, the spectral (Dirichlet) kernel $\Lambda_\infty(w)$ corresponding to the 'truncated periodogram' (where again $M = 40$) is plotted in Figure 3; in all three plots, w varies from $-2\pi/5$ to $+2\pi/5$.

Nevertheless, an immediate modification of $\hat{f}_h(w)$ can be constructed to yield a surely nonnegative spectral estimator. Define

$$\hat{f}_h^+(w) = \max(\hat{f}_h(w), 0), \tag{12}$$

i.e., $\hat{f}_h^+(w)$ is the positive part of $\hat{f}_h(w)$. Since $\hat{f}_h(w)$ is consistent, this modification obviously makes sense for large samples; if $f(w) > 0$, then with high probability $\hat{f}_h^+(w) = \hat{f}_h(w) > 0$, and if $f(w) = 0$, then it might as well be estimated by zero. In the following theorem it is shown that, even in finite samples, taking the positive part results in a better (with respect to MSE) estimator.

Theorem 2 *Let w be any point in $[-\pi, \pi]$. Then $MSE(\hat{f}_h^+(w)) \leq MSE(\hat{f}_h(w))$.*

It now follows that, under the assumptions of Theorem 1, the estimator $\hat{f}_h^+(w)$ is also consistent for $f(w)$, and has the desirable property of being nonnegative, in that case, the two estimators $\hat{f}_h^+(w)$ and $\hat{f}_h(w)$ are asymptotically equivalent as the following corollary of Theorem 2 states.

Corollary 1 *Let w be any point in $[-\pi, \pi]$. Then $E(\hat{f}_h^+(w) - \hat{f}_h(w))^2 \leq 2MSE(\hat{f}_h(w))$.*

The above equation is true for finite samples, but since $MSE(\hat{f}_h(w)) \to 0$ under regularity conditions, Corollary 1 can be expressed in asymptotic notation as $E(\hat{f}_h^+(w) - \hat{f}_h(w))^2 = O(MSE(\hat{f}_h(w)))$. Nevertheless, a sharper asymptotic result is also true under some assumptions, and will be proven later on, in our Section 4.

Although Theorem 2 does not explicitly quantify the bias of $\hat{f}_h^+(w)$, the following obvious corollary gives a a rough estimate which will be improved upon as well in Section 4.

Corollary 2 *Let w be any point in $[-\pi, \pi]$. Then $|Bias(\hat{f}_h^+(w))| \leq \sqrt{MSE(\hat{f}_h(w))}$.*

As a consequence of Theorem 1 and Corollary 2, by letting $M \sim aN^{1/(2r+1)}$, for some constant $a > 0$, the mean squared error of $\hat{f}_h^+(w)$ is of very small order, that is, $MSE(\hat{f}_h^+(w)) = O(N^{-2r/(2r+1)})$, and $|Bias(\hat{f}_h^+(w))| = O(N^{-r/(2r+1)})$. In particular, if f is smooth enough, (i.e., if r as defined in the assumptions of Theorem 1 is large), the bias of $\hat{f}_h^+(w)$ can be considered negligible even for moderately large M, and the rate of convergence of $\hat{f}_h^+(w)$ can be very close to $\sqrt{N}$.

2.4. The case of very smooth spectral densities and ARMA models. A comparison of the performance of $\hat{f}_h^+(w)$ and the performance of the usual nonparametric estimators possessing characteristic exponent equal to two is in order. Each of the latter results in an estimate with bias of order $O(1/M^l)$, where $l = \min(2, r)$, while $\hat{f}_h^+(w)$ has bias of order $o(1/M^r)$.

In other words, the fact that the characteristic exponent is finite and equal to 2 sets a ceiling on the bias-performance of the usual estimators, and does not allow the bias to become of smaller order, even if the true spectral density is known to be very smooth. It is worthwhile to note that for a large class of stationary time series, namely the class of Auto-Regressive processes with Moving Average residuals (ARMA), the spectral density f possesses any number of derivatives, (that is, r can be thought of as being infinite), and thus the bias of $\hat{f}_h^+(w)$ is negligible.

To see this, recall that the time series $\{X_t\}$ is said to follow an ARMA (n, m) model if it satisfies the difference equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_n X_{t-n} = Z_t + \psi_1 Z_{t-1} + \cdots + \psi_m Z_{t-m}, \tag{13}$$

for any integer t, where the Z_t's are uncorrelated random variables, with mean zero, and common variance σ^2. It is easy to show (cf., for example, Brockwell and Davis (1991)) that, provided the characteristic polynomial $1 - \phi_1 z - \cdots - \phi_n z^n = 0$ has no roots on the unit circle, the autocovariance $R(s)$ of $\{X_t\}$ decreases geometrically fast, i.e., $|R(s)| \leq De^{-d|s|}$, for some positive constants d and D, in which case $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$, for *any* positive integer r. It follows that, in the case the data are generated by such an ARMA model, M can be taken to increase very slowly with N, and the rate of convergence of $\hat{f}_h^+(w)$ will be $\sqrt{N/\log N}$ which is very close to $\sqrt{N}$, although the $\sqrt{N}$ rate is not exactly attainable.

Theorem 3 *Let w be any point in $[-\pi, \pi]$, and assume that the autocovariance $R(s)$ decreases geometrically fast, i.e., $|R(s)| \leq De^{-d|s|}$, for some positive constants d and D. Also assume that $\bar{m} \sim cM$, and that $M \sim A \log N$, where $c \in (0, 1)$ and $A \geq 1/(2cd)$ are some constants. Suppose in addition that the time series $\{X_t\}$ is such that the Bartlett spectral estimator $\tilde{f}(w)$ has a large-sample variance of order $O(M/N)$. Then, as $N \to \infty$,*

$$MSE(\hat{f}_h(w)) = O(\frac{\log N}{N}) \tag{14}$$

and

$$MSE(\hat{f}_h^+(w)) = O(\frac{\log N}{N}). \tag{15}$$

In the next section it will be shown that, in the special case where the time series $\{X_t\}$ can be thought to follow a Moving Average model of order m, i.e., if $\{X_t\}$ satisfies equation (13) with $\phi_k = 0, k \geq 1$, the rate of convergence of $\hat{f}_h^+(w)$ is actually *exactly* $\sqrt{N}$, and M can be taken to be a *fixed* number, not necessarily increasing to infinity.

2.5. The case of m-dependence and $\sqrt{N}$-consistency. Suppose now that the stationary time series $\{X_t\}$ is m-dependent, meaning that the set of random variables $\{X_t, t \leq 0\}$ is independent of the set of random variables $\{X_t, t > m\}$. Alternatively, just suppose that $R(s) = 0$, for all $|s| > m$, i.e., that the time series $\{X_t\}$ can be thought of as arising from a Moving Average (MA) model of order m, (cf. Brockwell and Davis (1991)). In both cases, the spectral density is given by the finite sum $f(w) = \frac{1}{2\pi}\sum_{s=-m}^{m} R(s)e^{-jsw}$, and it is quite obvious that the simple 'plug-in' estimate $\frac{1}{2\pi}\sum_{s=-m}^{m} \hat{R}(s)e^{-jsw}$ is a $\sqrt{N}$-consistent estimator of $f(w)$. Similarly, the 'truncated periodogram' $\frac{1}{2\pi}\sum_{s=-M}^{M} \hat{R}(s)e^{-jsw}$ is $\sqrt{N}$-consistent, *if M is fixed to a constant value greater or equal to m.*

However, a nonparametric spectral estimator with finite characteristic exponent will *not* estimate $f(w)$ at $\sqrt{N}$ rate of convergence, even if it is known that m-dependence holds. This loss of accuracy is of course due to the fact that if the lag-window is not exactly flat at the origin, there is a bias in the spectral estimator that can be made negligible only by letting M tend to infinity as N tends to infinity. On the other hand, since the variance of a nonparametric spectral estimator is generally proportional to M/N, it follows that the rate of convergence is $\sqrt{N/M} << \sqrt{N}$.

Recall that the truncated periodogram is just an extreme case (with $h = \infty$) of the estimator $\hat{f}_h(w)$. It would be quite interesting if $\hat{f}_h(w)$, and therefore, in view of Theorem 2, $\hat{f}_h^+(w)$ as well, share this desirable property of $\sqrt{N}$-consistency in the case $R(s) = 0$, for all $|s| > m$. This is in fact true, and is the subject of the next theorem.

Theorem 4 *Let w be any point in $[-\pi, \pi]$, and assume that $R(s) = 0$, for all $|s| > m$. Let $\bar{m}, M$ be* constants *satisfying $m \leq \bar{m} \leq M$. Suppose in addition that the time series $\{X_t\}$ is such that the Bartlett spectral estimator $\tilde{f}(w)$ has a large-sample variance of order $O(M/N)$.*

Then, as $N \to \infty$,

$$MSE(\hat{f}_h(w)) = O(1/N) \tag{16}$$

and

$$MSE(\hat{f}_h^+(w)) = O(1/N). \tag{17}$$

The point to be made here is that nonparametric spectral estimators given by $\hat{f}(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} k(s)\hat{R}(s)e^{-jsw}$, with lag-window $k(s) = 0$, for $|s| > M$, are essentially spectral densities of the Moving Average type of order M. Hence it might be reasonable to expect that the performance (bias, variance, rate of convergence) of $\hat{f}(w)$ should be significantly better in the special case the true spectral density $f(w)$ is itself of Moving Average type of order $m < M$, i.e., if the data arise from an MA(m) model. This is indeed true for the trapezoidal lag-window, although this is *not* the case for the more popular lag-windows possessing finite characteristic exponents. Note however that the trapezoidal lag-window was *not* tailor-made for m-dependent data; indeed its overall desirable properties were established in Theorem 1.

2.6. Choosing M in practice. Theorem 4 is extremely important for practical applications where the choice of the bandwidth parameter M is crucial. As mentioned in Section 2.2, the choice of h, or equivalently the choice of $\bar{m}$, is not nearly as important, and one might opt for the simple choice $h = 1$ and $\bar{m} = M/2$.

To choose M given the data $X_1, \ldots, X_N$, a practitioner will usually employ diagnostic tools, the prime of which is a correlogram, i.e., a plot of $\hat{R}(s)$; see Priestley (1981, p. 539). If it is observed that $\hat{R}(s) \simeq 0$, for all $|s| >$ some number $\hat{m}$, then it may be inferred that $\hat{m}$ is an estimate of m appearing in the assumptions of Theorem 4. It would then follow that $\bar{m}$ may be taken equal to $\hat{m}$, M may be taken equal to $\hat{m}(h+1)/h$, and the resulting estimators $\hat{f}_h(w)$ and $\hat{f}_h^+(w)$ would be very accurate, i.e., the error in the estimation of $f(w)$ by $\hat{f}_h(w)$ or by $\hat{f}_h^+(w)$ would be of very small order with high probability.

Note that this simple procedure for choosing M does *not* work well for the nonparametric spectral estimators possessing finite characteristic exponents. To see this, consider the simplest example where the time series $\{X_t\}$ is produced by the MA(1) model $X_t = Z_t + Z_{t-1}$, and the Z_t's are i.i.d. normal $N(0,1)$. Suppose that the sample size N is large enough so that from the correlogram it can easily be identified that $\hat{m} = 1$. From the above discussion it follows that $\hat{f}_1^+(w)$, with $\bar{m} = 1$ and $M = 2$, will be an accurate estimator of $f(w)$. However, it is apparent that estimating $f(w)$ by, say, a Bartlett estimator $\tilde{f}(w)$ with $M = 2$ will be highly inaccurate. In particular, since the sample is large enough to ensure that $\hat{R}(s) \simeq R(s)$, for a large range of s values, the absolute value of the systematic error in estimating $2\pi f(w)$ by $2\pi\tilde{f}(w)$ will be approximately equal to $|\cos w|/2M$, which can be made negligible only by taking M big enough, certainly much bigger than two.

2.7. Biased vs. unbiased autocovariances. It is interesting to note that the main reason the biased sample autocovariances $\hat{R}(s)$ are used in connection with a lag-window estimator is that the sequence $\hat{R}(\cdot)$ is nonnegative definite. This property implies that, if the spectral window is everywhere nonnegative, the resulting spectral estimator is necessarily nonnegative as well.

However, if one is *not* using a nonnegative spectral window, one might just as well use the unbiased sample autocovariances $\check{R}(s) = \hat{R}(s)N/(N-|s|)$; this is apparent considering that the sample autocovariances are heavily tapered anyway by the lag-window $\lambda(s)$, and the extra tapering by the factor $1 - |s|/N$ is negligible. Notably, using the unbiased sample autocovariances $\check{R}(s)$ instead of $\hat{R}(s)$ in definition (6) makes it possible to drop the condition $M^r/N \to 0$ from the assumptions of Theorem 1, substituting it with the more natural (and easier to satisfy) assumption $M/N \to 0$. Similarly, using the unbiased autocovariances $\check{R}(s)$ results in $\hat{f}_h(w)$ being *exactly* unbiased for *finite* samples under the assumptions of Theorem 4, i.e., $Bias(\hat{f}_h(w)) \equiv 0$, for any $w \in [-\pi, \pi]$, as long as $m \leq \bar{m} \leq M < N$.

3. Some finite-sample numerical results

Comparing equations (2) and (11) it follows that $Var(\hat{f}_h(w)) \sim \frac{3h+1}{h+1} Var(\tilde{f}(w))$; for example, if $h = 1$, it follows that $Var(\hat{f}_1(w)) \sim 2Var(\tilde{f}(w))$. Since $\frac{3h+1}{h+1}$ goes from 1 to 3 as h goes from 0 to ∞, this provides a further justification for the notion of the case $h = 1$ corresponding to the 'midpoint' between $h = 0$ (Bartlett) and $h = \infty$ (truncated periodogram).

The question may be asked, *"since $Var(\hat{f}_h(w)) > Var(\tilde{f}(w))$, how is it that $\hat{f}_h(w)$ has smaller MSE than the Bartlett?";* the answer lies with the choice of M. For a given sample size N, one would pick an M of the order of $N^{1/3}$ to use in conjunction with the Bartlett estimator, while the same researcher would pick an M of the order of $N^{1/(2r+1)}$ to use with $\hat{f}_h(w)$. In other words, $Var(\hat{f}_1(w)) = O(N^{1/(2r+1)}/N) = O(N^{-2r/(2r+1)})$; if $r > 1$, this is orders of magnitude less than the variance of the Bartlett estimator which is $O(N^{1/3}/N) = O(N^{-2/3})$. Arguably, the number of derivatives r will not be given in practice but, considering the simple example in Section 2.6, it is apparent that even a data-dependent choice would yield a much *smaller* M for use with $\hat{f}_h(w)$ than it would for use with the Bartlett estimator.

A further comparison is available if we define the integrated MSE (IMSE) of a spectral estimator $\hat{f}$ by the formula $IMSE(\hat{f}) = (1/2\pi)\int_0^{2\pi} MSE(\hat{f}(w))dw$. It is obvious that under the assumptions of Theorem 1, $IMSE(\hat{f}_h) = o(1/M^{2r}) + O(M/N)$, which is of smaller order of magnitude than the $IMSE$ of the Bartlett estimator $\tilde{f}$, provided M is chosen optimally in both cases, i.e., M proportional to $N^{1/(2r+1)}$ for $\hat{f}_h$, and M proportional to $N^{1/3}$ for $\tilde{f}$.

However, a criticism that might be raised is that the results offered so far are asymptotic in nature. To shed some light on the finite-sample performance of $\hat{f}_h(w)$, a simulation study was conducted. Using the statistical language S+, 190 independent realizations of a time series $\{X_t\}$ that obeys an ARMA (4,2) model of equation (13) with $\phi_1 = -1.352$, $\phi_2 = 1.338$, $\phi_3 = -0.662$, $\phi_4 = 0.240$, and $\psi_1 = -0.2$, $\psi_2 = 0.04$, were generated; each realization consisted of a stretch of 1000 observations, i.e., $N = 1000$ in this case.

The Bartlett estimator $\tilde{f}$ was computed for $M = 5, 10, 20, 30, 40, 50$ as an average of short periodograms; the Fast Fourier Transform (FFT) was employed throughout. Equation (5) was then used to compute $\hat{f}_h$ for the 15 possible different pairwise combinations of the available

Bartlett estimators. Note that the FFT computes a periodogram only for w of the form $2\pi k/M$, for $k = 0, 1, \ldots, M-1$, where $1/M$ is the bandwidth, and M is the block size. Since we needed to compare $\tilde{f}$ and $\hat{f}_h$ for different bandwidths, the technique of zero padding was used (cf. Gardner (1988)), and interpolations of all estimated spectra were computed on the grid $2\pi k/100$, for $k = 0, 1, \ldots, 99$; in all of our figures, w ranges from 0 to π, i.e., w of the form $2\pi k/100$, for $k = 0, 1, \ldots, 49$.

To get a feeling about what kind of information would be available to a practitioner faced with a stretch of 1000 observations from series $\{X_t\}$, one of the 190 realizations was selected at random, and some of the computed spectral estimators $\tilde{f}$ and $\hat{f}_h$ with different bandwidths are presented in Figures 4 and 5 respectively; the true spectral density f is also shown for comparison. Just from looking at the pictures, it is apparent that the Bartlett estimator $\tilde{f}$ with $M = 20$ is the best among the different Bartlett estimators, and that the estimator $\hat{f}_h$ with $M = 10$ and $\bar{m} = 5$ is best among the different bias-corrected estimators, although the one with $M = 20$ and $\bar{m} = 5$ is a close competitor; this intuitive observation is in fact true and will be confirmed by the MSE calculations that follow. Similarly, it is quite obvious that the Bartlett estimator $\tilde{f}$ with $M = 5$ is quite oversmoothed, while $\tilde{f}$ with $M = 50$ is undersmoothed, both extremes being suboptimal. In terms of the bias-corrected estimators $\hat{f}_h$, it seems that the worst among them is the one with $M = 50$ and $\bar{m} = 40$, both because its corresponding bandwidth is too small, as well as because its trapezoidal window more closely approximates the Dirichlet than any other window used in this simulation, i.e., it is the one with highest h; its bad performance can be appreciated just from the fact that it goes negative near $w = 0.88$ which is the mode of the true f.

In addition, a sample correlogram was also computed from the same chosen realization, and displayed in Figure 6 together with the true correlogram. Note that the true correlogram tapers off exponentially for large lags, and $R(s) \simeq 0$, for $s > 6$; similarly, despite the spurious waves displayed by $\hat{R}(s)$ for large lags, an experienced data analyst might infer that really $R(s) \simeq 0$, for $s >$ some $\hat{m}$, where $\hat{m} = 6$ or 7. Hence, following the rationale of Section 2.6, he/she would choose a $\bar{m}$ of about 6 or 7; thus, the previously mentioned observation that the bias-corrected estimators with $\bar{m} = 10$ or 5 are the 'best' reinforces the validity of the recommendations in

Section 2.6.

Nevertheless, the main point of the simulation is that the expected value of a spectral estimator (which is required to calculate its bias) can now be approximated by the empirical average of the 190 independent spectral estimators that are available. However, because of the zero padding employed, the apparent variability of the 190 independent spectral estimators was reduced, rendering empirical variances unreliable. For this reason, equations (2) and (11) were used throughout, whenever variance estimates were needed, e.g., for the computation of the MSE of estimators. Note though, that in a case where no zero padding was used, the empirical variance of the Bartlett estimator $\tilde{f}$ with $M = 50$ was found to be very close to the asymptotic equation (2); see Figure 7. This observation leads us to believe that the asymptotic equations (2) and (11) are reasonably good approximations for a sample of this size. It should be stressed however, that the asymptotic equations (2) and (11) would *not* be readily available to the practitioner, because he/she would not know the true value of f; the practitioner would likely substitute an estimator in place of the unknown f, and the variance estimator would consequently suffer from all the inaccuracies inherent in spectral estimation.

In Figure 8, the bias and (asymptotic) standard deviation of Bartlett estimators with different bandwidths are compared; as expected, increasing M reduces bias and increases the standard deviation so the different curves shown are 'nested' within one another. Similarly, in Figures 9 to 12, the bias and (asymptotic) standard deviation of some of the available bias-corrected estimators with different combinations of $\bar{m}$ and M are shown; comparing these bias-standard deviation curves for a fixed w near $w = 0.88$ where the true f has its mode, it is apparent that the estimator $\hat{f}_h$ with M=10 and $\bar{m}$=5 is the best in the MSE sense among the ones shown.

In order to provide a more thorough comparison of pointwise performance, the MSE of *all* available estimators is shown in Figures 13 to 17 as a function of w. Figure 13 contains the MSE of Bartlett estimators from which it follows that the choice $M = 20$, which is pictured in both Figures 13 (a) and (b), is optimal for any chosen w. Figures 14 to 17 present the MSE performance of the bias-corrected estimator $\hat{f}_h$ for different combinations of M and $\bar{m}$; the grouping of curves to the figures was done with the purpose of having the curves in each figure nested within one another, so that the figures are easily interpretable. Is is easy to see that, except in a neighborhood of $w = 0.5$, the estimator $\hat{f}_h$ with M=10 and $\bar{m}$=5 dominates all the available estimators in terms of their MSE. Even in the the neighborhood of $w = 0.5$, the MSE of $\hat{f}_h$ with M=10 and $\bar{m}$=5 is comparable to the MSE of its closest competitors, therefore allowing us to conclude that it is globally the best.

To have a further confirmation of the above conclusion, the integrated MSE (IMSE) of all available estimators was computed and presented in Table 1 as a global measure of performance. Not surprisingly, it follows from Table 1 that the best (in an IMSE sense) estimator is $\hat{f}_h$ with M=10 and $\bar{m}$=5, i.e., $h = 1$, followed by $\hat{f}_h$ with M=20 and $\bar{m}$=5, i.e., $h = 1/3$, in turn followed by the Bartlett estimator $\tilde{f}$ with $M = 20$.

The information presented in Table 1 serves to confirm that there is a definite finite-sample advantage in employing the proposed bias-correction. In particular, the MSE of an optimized bias-corrected estimator (in this case: $\hat{f}_h$ with M=10 and $\bar{m}$=5) can be significantly less than the MSE of an optimized Bartlett estimator (in this case: the one with M=20); in our simulation the MSE was reduced by more than 25%. This reduction in MSE is essentially achieved because *the optimized bias-corrected estimator requires a smaller M (larger bandwidth) as compared to the optimized Bartlett estimator!*

Similarly, Table 1 exemplifies the fact that the choice of the design parameters M and $\bar{m}$ is still crucial; however, it also provides some empirical support for our recommendations of Section 2.6 regarding these choices. More specifically, choosing $\bar{m} \simeq \hat{m}$, where $\hat{m}$ is the smallest integer such that $\hat{R}(s) \simeq 0$ for $s > \hat{m}$ seems to work. Having chosen $\bar{m}$, it remains to choose M, or equivalently, h. From Table 1 it is apparent that choosing $h = 1$ seems advisable, thus indicating the choice $M \simeq 2\hat{m}$. Using an h that lies between 0 and 1 might be an acceptable choice as well, with a subsequent choice of $M \simeq \hat{m}(h+1)/h$.

Since all our discussion so far pertained to the estimator $\hat{f}_h$, one might inquire about the MSE and IMSE performance of $\hat{f}_h^+$. As it turns out, the empirical IMSE's of $\hat{f}_h^+$ were generally found to be a bit smaller than the corresponding IMSE's for $\hat{f}_h$. However, the difference is so small that the IMSE table of $\hat{f}_h^+$ for different combinations of M and $\bar{m}$ agrees with Table 1 to the first three or four decimal points; thus it is not given here, especially because, since the entries to the Table are empirical, one would not trust them to be accurate to more than the two decimals shown. However, it is interesting to ask how often $\hat{f}_h^+$ was identical to $\hat{f}_h$ in our simulation, i.e., how often $\hat{f}_h$ yielded negative estimates. For this purpose, Tables 2 and 3 are presented that have as entries the empirically estimated probabilities $Prob\{\hat{f}_h(0.88) < 0\}$, and $Prob\{\hat{f}_h(\pi) < 0\}$ respectively, for different combinations of $\bar{m}$ and M; note that $w = 0.88$ is the mode of f, while $w = \pi$ is where f achieves its minimum.

From Table 2 it is apparent that only the very bad estimators, that possess both too high M's as well as too high an h parameter, show any occurence of $\hat{f}_h(0.88) < 0$. On the other hand, Table 3 shows that occurences of $\hat{f}_h(\pi) < 0$ are quite frequent, even in our best estimator $\hat{f}_h$ with $\bar{m} = 5$ and $M = 10$. Note however, that estimating $f(\pi) \simeq 0.073$ by 0 would actually be most accurate; therefore, it is exactly in this case where $f(\pi)$ is of very small magnitude and $\hat{f}_h$ has a good chance of being negative that the estimator $\hat{f}_h^+(\pi)$ would present a significant improvement over $\hat{f}_h(\pi)$. To explain the fact that the IMSE of $\hat{f}_h^+$ is the same to that of $\hat{f}_h$ (both with $\bar{m} = 5$ and $M = 10$), note that the IMSE is dominated by the large MSE's occuring near the mode $w = 0.88$. The MSE of $\hat{f}_h^+(\pi)$ is significantly less than the MSE of $\hat{f}_h(\pi)$, but the latter is extremely small to start with; in other words, the relative reduction in MSE might be significant, while the the absolute reduction (the difference of the two MSE's) is *not*.

	$M=5$	$M=10$	$M=20$	$M=30$	$M=40$	$M=50$
$\bar{m} = 0$	2.85	0.83	0.45	0.50	0.59	0.71
$\bar{m} = 5$		0.33	0.41	0.54	0.67	0.81
$\bar{m} = 10$			0.54	0.68	0.81	0.95
$\bar{m} = 20$				0.95	1.08	1.22
$\bar{m} = 30$					1.37	1.50
$\bar{m} = 40$						1.85

Table 1. Entries are the empirically estimated integrated MSE (IMSE) of $\hat{f}_h$ for different combinations of $\bar{m}$ and M; the case $\bar{m} = 0$ corresponds to a pure Bartlett estimator with parameter M.

	$M=5$	$M=10$	$M=20$	$M=30$	$M=40$	$M=50$
$\bar{m}=5$		0.000	0.000	0.000	0.000	0.000
$\bar{m}=10$			0.000	0.000	0.000	0.000
$\bar{m}=20$				0.000	0.000	0.000
$\bar{m}=30$					0.026	0.005
$\bar{m}=40$						0.105

Table 2. Entries are the empirically estimated probabilities $Prob\{\hat{f}_h(0.88) < 0\}$, for $\hat{f}_h$ resulting from different combinations of $\bar{m}$ and M.

	$M=5$	$M=10$	$M=20$	$M=30$	$M=40$	$M=50$
$\bar{m}=5$		0.368	0.105	0.042	0.021	0.005
$\bar{m}=10$			0.121	0.058	0.032	0.011
$\bar{m}=20$				0.242	0.080	0.058
$\bar{m}=30$					0.310	0.163
$\bar{m}=40$						0.337

Table 3. Entries are the empirically estimated probabilities $Prob\{\hat{f}_h(\pi) < 0\}$, for $\hat{f}_h$ resulting from different combinations of $\bar{m}$ and M.

4. Practical comments and conclusions

4.1. Some practical comments. In Sections 2 and 3, an intuitive proposal for bias-corrected nonparametric spectral estimators was introduced and analyzed, and it was shown that it essentially reduces to taking the positive part of a spectral estimator with trapezoidal lag-window. It was also shown that the proposed estimator can be easily computed as (the positive part of) a linear combination of two Bartlett estimators with different bandwidths. However, the presented bias reduction methodology is not limited to the example studied here in detail; indeed, a general method was introduced to combine two function estimators in order to obtain a third estimator with smaller bias.

To focus on a specific important application of the proposed bias reduction scheme, consider the case in which the objective is estimation of $Var(\sqrt{N}\bar{X}_N)$, where $\bar{X}_N = N^{-1}\sum_{i=1}^{N} X_i$ is the sample mean. It is easy to see that $2\pi\tilde{f}(0)$, which is a constant multiple of the Bartlett spectral estimator evaluated at point 0, is a consistent estimator of $Var(\sqrt{N}\bar{X}_N)$, and therefore, $\sqrt{2\pi\tilde{f}(0)}$ is a consistent estimator of the standard error $\sqrt{Var(\sqrt{N}\bar{X}_N)}$. As a matter of fact, the estimator $2\pi\tilde{f}(0)$ comes up very naturally as the resampling ('moving blocks' bootstrap) and subsampling ('moving blocks' jackknife) variance estimator ((cf. Künsch (1989), Liu and Singh (1992), Politis and Romano (1992a,c)); it also comes up as the 'batch means' variance estimator in the steady state simulation literature (cf. Meketon and Schmeiser (1984), Welch (1987), Song and Schmeiser (1988, 1992)). The bias reduction methodology developed in Sections 2 and 3, can be used to combine two such estimators (with different block-batch sizes) to obtain a more accurate variance estimate.

Regarding the important problem of setting confidence intervals for $f(w)$ on the basis of $\hat{f}_h(w)$ or $\hat{f}_h^+(w)$ there are two avenues, one based on a central limit theorem, and the other using resampling and subsampling methods; for more details on the second approach, see Politis and Romano (1992a,c), and Politis, Romano, and Lai (1992)). To elaborate on the first method, note that $\hat{f}_h(w)$ will be asymptotically normal under regularity conditions (cf. Hannan (1970), Brillinger (1975), Rosenblatt (1984, 1985)), and so will $\hat{f}_h^+(w)$ as the following theorem demonstrates.

Theorem 5 *Under the conditions of Theorem 1, the additional condition that* $N^{1/(2r+1)} = O(M)$*, as well as the assumption that* $\sqrt{N/M}(\hat{f}_h(w) - f(w))$ *is asymptotically normal, the following are true:*
(a) If $f(w) = 0$*, the large-sample distribution of either* $\sqrt{N/M}\hat{f}_h(w)$ *or* $\sqrt{N/M}\hat{f}_h^+(w)$ *degenerates to a point mass at the origin as* $N \to \infty$*;*
(b) If $f(w) > 0$*, then* $\sqrt{N/M}(\hat{f}_h(w) - f(w))$ *and* $\sqrt{N/M}(\hat{f}_h^+(w) - f(w))$ *have the same asymptotic normal* $N(0, \sigma^2)$ *distribution, where* $\sigma^2 = \frac{3h+1}{h+1}\frac{2}{3}f^2(w)(1 + \eta(w))$.

Note that choosing M big enough to ensure that $N^{1/(2r+1)} = O(M)$ ensures that the limiting normal distributions of Theorem 5 have zero mean, and thus probability statements using the approximate normal distributions can be inverted to yield confidence intervals for $f(w)$. In other words, the choice of M should be such that $Bias(\hat{f}_h(w)) = o(\sqrt{M/N})$, $Var(\hat{f}_h(w)) = O(M/N)$, and $MSE(\hat{f}_h(w)) = O(M/N)$ as well.

The two cases where $f(w) > 0$ and $f(w) = 0$ can actually be included in a single formulation, namely that $\sqrt{N/M}(\hat{f}_h(w) - f(w))$ and $\sqrt{N/M}(\hat{f}_h^+(w) - f(w))$ have the same asymptotic normal $N(0, \sigma^2)$ distribution, and allowing for the fact that $\sigma^2 = 0$ if $f(w) = 0$. However, in the rather more interesting second case where $f(w) > 0$, much more can be said regarding the closeness of $\hat{f}_h^+(w)$ to $\hat{f}_h(w)$. The following lemma sharpens some results of Section 2.3 in that case; it is proven under conditions similar to the conditions of Theorem 5, albeit a bit more general, allowing for the possibility that there is a remaining bias in the limiting normal distribution.

Lemma 2 *Let* w *be any point in* $[-\pi, \pi]$ *such that* $f(w) > 0$*. Assume that as* $N \to \infty$*,* $M \to \infty$ *but* $M/N \to 0$*,* $\sqrt{N/M}Bias(\hat{f}_h(w)) \to C_b$*,* $(N/M)Var(\hat{f}_h(w)) \to C_v^2 > 0$*, and that* $\sqrt{N/M}(\hat{f}_h(w) - f(w))$ *has an asymptotic normal* $N(C_b, C_v^2)$ *distribution; here* C_b *and* C_v^2 *are constants that may depend on* f *and* w*. Then* $E(\hat{f}_h^+(w) - \hat{f}_h(w))^2 = o(MSE(\hat{f}_h(w)))$.

Lemma 2 has the following interesting corollary.

Corollary 3 *Under the assumptions of Lemma 2 the following are true:*
(a) $Bias(\hat{f}_h^+(w)) = Bias(\hat{f}_h(w)) + o(\sqrt{M/N})$*, and*
(b) $Var(\hat{f}_h^+(w)) = Var(\hat{f}_h(w)) + o(M/N)$.

4.2. Conclusions. Spectral density estimation is now an almost fifty year old field, and it would seem at first that whatever could be said about it has already been said, and one should be able to look it up in a textbook, say Priestley's (1981) comprehensive treatise. Nevertheless, this premise is not necessarily true; to make the point we will now compare the results of the present paper to its closest relative, namely the general theory of Parzen (1957a,b).

To start, note that in Parzen's (1957a,b) pioneering development it was proved that, if the characteristic exponent of a lag-window is equal to q, then the bias of the corresponding spectral estimator is of order $O(1/M^p)$, where $p = \min(q, r)$, and r was defined in the assumptions of Theorem 1. It is apparent that the order $O(1/M^r)$ for the bias will be achieved *only if* the practitioner happens to choose a lag-window with $q \geq r$; since r is not given in any practical application, choosing q for each data-set becomes a very hard issue.

In addition, the possible nonpositivity of estimators corresponding to lag-windows with large q was considered to be a major drawback, sufficient to limit consideration to lag-windows with characteristic exponent not greater than two. In Section 2.3 it was shown how this nonpositivity is easily side-stepped without sacrificing the good MSE performance of the estimator.

In the same vein, the only estimator with infinite characteristic exponent considered in Parzen (1957a) was the truncated periodogram, which is well known to possess undesirable properties (cf. Hannan(1970)). Observe that the spectral window $\Lambda_\infty(w)$ of the truncated periodogram (see Figure 3) exhibits quite prominent positive and negative side-lobes which may introduce spurious details in the estimate of a spectral density containing sharp peaks. It is important to point out that the spectral window $\Lambda_1(w)$ of the "$2f - f$" trapezoidal rule (see Figure 2) does not exhibit such behaviour.

Note also that Parzen (1957a) introduced the family of lag-windows given by

$$k(s) = \begin{cases} 1 - |s/M|^q & \text{for } |s| \leq M \\ 0 & \text{for } |s| > M, \end{cases}$$

parameterized by the characteristic exponent q. Similarly to the family of trapezoidal lag-windows, Parzen's family has the Bartlett estimator and the truncated periodogram at its extreme points ($q = 1$ and $q = \infty$). Nonetheless, the two families of lag-windows are remarkably different; in particular, *all* lag-windows in the trapezoidal family for $h \in (0, \infty)$ have an infinite characteristic exponent, and share the same properties that are summarized below.

- Both $\hat{f}_h(w)$ and $\hat{f}_h^+(w)$ can be computed easily and fast, taking into account that the actual computation of Bartlett's estimator as an average of short periodograms is extremely fast (cf. Gardner (1988)).
- The MSE of $\hat{f}_h(w)$ and $\hat{f}_h^+(w)$ can be of very small order provided the function $f(w)$ is smooth, having a number of derivatives.
- The rate of convergence of $\hat{f}_h(w)$ and $\hat{f}_h^+(w)$ is $\sqrt{N/\log N}$ if the data are generated by an ARMA model, and $\sqrt{N}$ if the data are MA(m) or m-dependent.
- Last but not least in importance is that working with $\hat{f}_h(w)$ and/or $\hat{f}_h^+(w)$ significantly simplifies the difficult problem of choosing the bandwidth of the spectral estimator in practice, at least in the case where the sample autocovariances seem to be negligible from some point on.

Acknowledgement

Helpful conversations on the problem of bias with Professors E. Parzen (Texas A&M University) and M. Rosenblatt (University of California, San Diego) are gratefully acknowledged.

Appendix: Technical proofs.

PROOF OF LEMMA 1. Parts (a)–(c) follow immediately from representation (9) and the properties of the Fejér kernel (cf. Brockwell and Davis (1991)). The proof of part (d) can be found in Politis and Romano (1992b).

PROOF OF THEOREM 1. Let w be any point in $[-\pi,\pi]$, and observe (cf. Parzen (1957a), Priestley (1981, p. 459)) that

$$Bias(\hat{f}_h(w)) \equiv E\hat{f}_h(w) - f(w) = A_1 + A_2 + A_3$$

where

$$A_1 = \frac{1}{2\pi}\sum_{s=-N+1}^{N-1} (\lambda(s)-1)\,R(s)e^{-jsw}$$

$$A_2 = -\frac{1}{2\pi N}\sum_{s=-N+1}^{N-1} |s|\lambda(s)R(s)e^{-jsw}$$

$$A_3 = -\frac{1}{2\pi}\sum_{|s|\geq N} R(s)e^{-jsw}.$$

But $|A_3| \leq \frac{1}{2\pi}\sum_{|s|\geq N}|R(s)| \leq \frac{1}{2\pi N^r}\sum_{|s|\geq N}|s|^r|R(s)| = o(1/N^r)$, since $\sum|s|^r|R(s)| < \infty$. Similarly, $|A_2| = O(1/N)$, using the facts $|\lambda(s)| \leq 1$, and $\sum|s||R(s)| < \infty$.

To complete the proof of equation (10), the term A_1 will now be shown to be of order $o(1/M^r)$. Note that A_1 can be split into three terms, $A_1 = a_1 + a_2 + a_3$, where

$$a_1 = \frac{1}{2\pi}\sum_{|s|\leq\bar{m}} (\lambda(s)-1)\,R(s)e^{-jsw}$$

$$a_2 = \frac{1}{2\pi}\sum_{\bar{m}<|s|\leq M} (\lambda(s)-1)\,R(s)e^{-jsw}$$

$$a_3 = \frac{1}{2\pi}\sum_{M<|s|<N} (\lambda(s)-1)\,R(s)e^{-jsw}.$$

First observe that $a_1 = 0$, because $\lambda(s) = 1$ for $|s| \leq \bar{m}$. Now

$$|a_2| \leq \frac{1}{\pi}\sum_{\bar{m}<s\leq M} |\lambda(s)-1||R(s)|.$$

But $\lambda(s) = 1 - \frac{s-\bar{m}}{M-\bar{m}}$ for $\bar{m} < s \leq M$. Thus,

$$|a_2| \leq \frac{1}{\pi} \sum_{\bar{m} < s \leq M} \frac{s-\bar{m}}{M-\bar{m}} |R(s)|.$$

It is obvious that if $r = 1$, then $a_2 = o(1/M)$. On the other hand, if $r > 1$, we have

$$|a_2| \leq \frac{1}{\pi \bar{m}^{r-1}} \sum_{\bar{m} < s \leq M} s^{r-1} \frac{s-\bar{m}}{M-\bar{m}} |R(s)| = o(1/M^r),$$

where it was used that $\sum |s|^r |R(s)| < \infty$, and that both $\bar{m}$ and $M - \bar{m}$ are asymptotically proportional to M. By a similar argument it is also shown that $a_3 = o(1/M^r)$ as well; since all the bounds above do not depend on w, equation (10) follows.

Now by computing the Hilbert l_2 norm of the lag-window $\lambda(\cdot)$, equation (11) follows, and the theorem is proven.

PROOF OF THEOREM 2. The proof of Theorem 2 is a consequence of the following general proposition; note that for the proof of the proposition, no assumptions whatsoever are required (independence, stationarity, etc.) regarding the probability structure of the sample.

Proposition 1 *Let $\theta \geq 0$ be an unknown parameter, and let $\hat{\theta}_N$ be an estimator of θ based on a sample of size N. Then, $MSE(\hat{\theta}_N^+) \leq MSE(\hat{\theta}_N)$, where $\hat{\theta}_N^+ \equiv \max(\hat{\theta}_N, 0)$.*

PROOF OF PROPOSITION 1. Note that

$$|\hat{\theta}_N^+ - \theta| \leq |\hat{\theta}_N - \theta| \tag{18}$$

always. Indeed, either $\hat{\theta}_N^+ = \hat{\theta}_N$ and equality holds in the above, or $\hat{\theta}_N < \hat{\theta}_N^+ = 0$, in which case $|\hat{\theta}_N^+ - \theta| < |\hat{\theta}_N - \theta|$. Squaring and taking expectations in equation (18) proves the proposition.

Theorem 2 now follows from Proposition 1 by making the obvious identification $\theta = f(w)$, $\hat{\theta}_N = \hat{f}_h(w)$, and $\hat{\theta}_N^+ = \hat{f}_h^+(w)$.

PROOF OF THEOREM 3. Recall from the proof of Theorem 1 that

$$Bias(\hat{f}_h(w)) = A_1 + A_2 + A_3$$

where A_2 is of order $O(1/N)$; under the extra assumptions we have $A_3 = O(e^{-dN})$ now. Since $a_1 = 0$, we just need to consider the terms a_2 and a_3; by a simple calculation it follows that both a_2 and a_3 are of order $O(e^{-d\bar{m}})$.

Thus the $Bias(\hat{f}_h(w)) = O(e^{-cdM})$, uniformly in $w \in [-\pi, \pi]$, where $c = \bar{m}/M$, and it is easy to show that the $MSE(\hat{f}_h(w))$ is asymptotically minimized by letting $M \sim A \log N$, for some $A \geq 1/(2cd)$. By this choice of M, $Bias(\hat{f}_h(w)) = O(1/\sqrt{N})$, $Var(\hat{f}_h(w)) = O(\log N/N)$, and $MSE(\hat{f}_h(w)) = O(\log N/N)$. Finally, Theorem 2 implies that $MSE(\hat{f}_h^+(w)) = O(\log N/N)$ as well.

PROOF OF THEOREM 4. Note that since $N \to \infty$, we can assume without loss of generality that $N > M$. As in the proof of Theorem 3, we only need to consider the terms A_3, a_2, and a_3, in the decomposition of the bias; however now $A_3 = 0$, since it is assumed that $R(s) = 0$, for $|s| > m$. The term A_1 now can be written as

$$A_1 = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} (\lambda(s) - 1) R(s) e^{-jsw} = \frac{1}{2\pi} \sum_{s=-m}^{m} (\lambda(s) - 1) R(s) e^{-jsw},$$

where again it was used that $R(s) = 0$, for $|s| > m$. But $\lambda(s) = 1$ for $|s| \leq \bar{m}$, and since it is assumed that $\bar{m} \geq m$, it follows that $A_1 = 0$.

Putting it all together, it is seen that $Bias(\hat{f}_h(w)) = O(1/N)$, uniformly in $w \in [-\pi, \pi]$. Now $Var(\hat{f}_h(w)) = O(M/N) = O(1/N)$ by Theorem 1 and the assumption that M is a constant. Hence, equation (16) follows. To complete the proof of the theorem, note that equation (17) follows from equation (16) and Theorem 2.

PROOF OF THEOREM 5. Since $MSE(\hat{f}_h^+(w)) \leq MSE(\hat{f}_h(w))$ by Theorem 2 and $MSE(\hat{f}_h(w)) = o(M/N)$ by our assumptions, part (a) follows.

Now note that, since $f(w) > 0$, an application of Chebychev's inequality yields

$$Prob\{\hat{f}_h^+(w) \neq \hat{f}_h(w)\} = Prob\{\hat{f}_h(w) < 0\} \leq \frac{MSE(\hat{f}_h(w))}{f^2(w)} \to 0,$$

as $N \to \infty$. So the estimators $\hat{f}_h^+(w)$ and $\hat{f}_h(w)$ are identical with probability tending to one; thus they have the same limit distribution, and part (b) follows.

PROOF OF LEMMA 2. The proof of Lemma 2 is a consequence of the following general proposition, by making the identifications $\tau_N = \sqrt{N/M}$, $\theta = f(w)$, $\hat{\theta}_N = \hat{f}_h(w)$, and $\hat{\theta}_N^+ = \hat{f}_h^+(w)$.

Proposition 2 *Let $\theta > 0$ be an unknown parameter, let $\hat{\theta}_N$ be an estimator of θ based on a sample of size N, and let $\hat{\theta}_N^+ \equiv \max(\hat{\theta}_N, 0)$. Assume that τ_N is a sequence such that $\tau_N \to \infty$ as $N \to \infty$, $E\left(\tau_N(\hat{\theta}_N - \theta)\right) \to b(\theta) \geq 0$, and $Var\left(\tau_N(\hat{\theta}_N - \theta)\right) \to \sigma^2(\theta) > 0$, as $N \to \infty$, and that $\tau_N(\hat{\theta}_N - \theta)$ has an asymptotic normal $N(b(\theta), \sigma^2(\theta))$ distribution. Then, $\tau_N^2 E(\hat{\theta}_N^+ - \hat{\theta}_N)^2 \to 0$, as $N \to \infty$.*

PROOF OF PROPOSITION 2. Let $Z_N = \tau_N(\hat{\theta}_N - \theta)$, and note that, by the asymptotic normality and the convergence of the first two moments, $\{Z_N^2\}$ is uniformly integrable; cf., for example, Serfling (1980, Lemma B, p. 15).

We want to show that, for any $\epsilon > 0$, there is a N_0 such that $\tau_N^2 E(\hat{\theta}_N^+ - \hat{\theta}_N)^2 < \epsilon$, for all $N > N_0$. So fix $\epsilon > 0$; by the uniform integrability of $\{Z_N^2\}$ it follows that there is a $c > 0$ such that $\sup_N E\left(Z_N^2 1\{|Z_N| > c\}\right) < \epsilon$, where $1\{B\}$ is the indicator of set B.

Now let N_0 be the smallest integer such that $\tau_N \theta > c$ for all $N > N_0$, and observe that

$\tau_N^2 E(\hat{\theta}_N^+ - \hat{\theta}_N)^2 = \tau_N^2 E\left(\hat{\theta}_N^2 1\{\hat{\theta}_N < 0\}\right)$

$\leq E\left(Z_N^2 1\{Z_N < \tau_N \theta\}\right)$ (because if $\hat{\theta}_N < 0$, then $|\hat{\theta}_N - 0| \leq |\hat{\theta}_N - \theta|$)

$\leq E\left(Z_N^2 1\{Z_N < -c\}\right)$ (for all $N > N_0$)

$\leq E\left(Z_N^2 1\{|Z_N| > c\}\right) < \epsilon$ (by construction).

Hence the proposition is proven, and so is Lemma 2.

References

[1] Bartlett,M.S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *J.Roy.Statist.Soc.Suppl.*, 8, 27-41.

[2] Blackman, R.B. and Tukey, J.W.(1959), *The measurement of Power Spectra from the point of view of Communication Engineering*, Dover, New York.

[3] Brillinger, D.R. (1975), *Time Series, Data Analysis and Theory*, Holt,Rinehart and Winston, New York.

[4] Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods, 2nd ed.*, Springer, New York.

[5] Devroye, L. (1987), *A course in density estimation*, Birkhäuser, Boston.

[6] Gardner, W. A. (1988), *Statistical Spectral Analysis*, Prentice Hall, New Jersey.

[7] Grenander,U. and Rosenblatt,M.(1957), *Statistical Analysis of Stationary Time Series*, Dover, New York.

[8] Hannan, T. (1970), *Multiple Time Series*, John Wiley, New York.

[9] Katznelson, Y. (1968), *An Introduction to Harmonic Analysis*, Dover, New York.

[10] Künsch, H. (1989), The jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, 17, 1217-1241.

[11] Liu, R.Y. and Singh, K. (1992), Moving Blocks Jackknife and Bootstrap Capture Weak Dependence, in *Exploring the Limits of Bootstrap*, (edited by Raoul LePage and Lynne Billard), John Wiley, pp. 225-248.

[12] Meketon, M.S., and Schmeiser, B. (1984), Overlapping batch means: something for nothing? *Proc. Winter Simulation Conference*, S. Sheppard, U. Pooch, and D. Pegden, (eds.), pp. 227-230.

[13] Parzen, E. (1957a), On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.,* 28, 329-348.

[14] Parzen, E. (1957b), Choosing an estimate of the spectral density function of a stationary time series. *Ann. Math. Statist.,* 28, 921-932.

[15] Politis,D.N., and Romano,J.P. (1992a), A General Resampling Scheme for Triangular Arrays of α-mixing Random Variables with Application to the Problem of Spectral Density Estimation, *Ann. Statist.*, vol. 20, No. 4, December 1992, pp. 1985-2007.

[16] Politis,D.N., and Romano,J.P. (1992b), Bias-corrected nonparametric spectral estimation, Technical Report No. 92-50, Department of Statistics, Purdue University.

[17] Politis,D.N., and Romano,J.P. (1992c), A General Theory for Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions, Technical Report No. 399, Department of Statistics, Stanford University.

[18] Politis,D.N., Romano,J.P., and Lai, T.-L. (1992), Bootstrap Confidence Bands for Spectra and Cross-Spectra, *IEEE Trans. Signal Proc.*, vol. 40, No. 5, 1206-1215.

[19] Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press.

[20] Quenouille, M. (1949), Approximate tests of correlation in time series, *J. Royal Statist. Soc., B*, 11, 68-84.

[21] Rosenblatt, M. (1984), Asymptotic normality, strong mixing and spectral density estimates, *Ann. Prob.*, vol. 12, No. 4, pp. 1167-1180.

[22] Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhäuser, Boston.

[23] Serfling, R.J. (1980), *Approximation theorems of mathematical statistics*, John Wiley, New York.

[24] Song, W. and Schmeiser, B. (1988), Minimal-MSE linear combinations of variance estimators of the sample mean, *Proc. Winter Simulation Conference*, M. Abrams, P. Haigh, and J. Comfort (eds.), pp. 414-421.

[25] Song, W. and Schmeiser, B. (1992), Variance of the sample mean: properties and graphs of quadratic-form estimators, Manuscript submitted to *Operations Research.*

[26] Welch, P.D. (1987), On the relationship between batch means, overlapping batch means and spectral estimation, *Proc. Winter Simulation Conference*, A. Thesen, H. Grant, and W.D. Kelton, (eds.), pp. 320-323.

[27] Zhurbenko, I.G. (1986), *The Spectral Analysis of Time Series*, North-Holland, Amsterdam.

CAPTIONS FOR FIGURES.

FIGURE 1. The spectral (Fejér) kernel $\Lambda_0(w)$; case $M = 40$.

FIGURE 2. The spectral kernel $\Lambda_1(w)$; case $M = 40$.

FIGURE 3. The spectral (Dirichlet) kernel $\Lambda_\infty(w)$; case $M = 40$.

FIGURE 4. Bartlett estimators with different bandwidths (dotted line) vs. the true spectral density (solid line):

(a) M= 5;

(b) M=10;

(c) M=20;

(d) M=30;

(e) M=40;

(f) M=50.

FIGURE 5. Bias-corrected estimators $\hat{f}_h(w)$ (dotted line) vs. the true spectral density (solid line):

(a) M= 50 and $\bar{m}$= 5;

(b) M= 30 and $\bar{m}$= 5;

(c) M= 20 and $\bar{m}$= 5;

(d) M= 10 and $\bar{m}$= 5;

(e) M= 40 and $\bar{m}$= 10;

(f) M= 20 and $\bar{m}$= 10;

(g) M= 50 and $\bar{m}$= 20;

(h) M= 30 and $\bar{m}$= 20;

(i) M= 50 and $\bar{m}$= 30;

(j) M= 50 and $\bar{m}$= 40.

FIGURE 6. A sample correlogram (dotted line) vs. the true correlogram (solid line).

FIGURE 7. Empirical variance of the Bartlett $M = 50$ estimator (dotted line) vs. the variance given by the asymptotic formula (solid line).

FIGURE 8. Bias and (asymptotic) standard deviation of the Bartlett estimators with $M = 5, 10, 20, 30, 40$ shown as functions of $w \in [0, \pi]$; the curves are nested within one another:
(a) Smallest -in magnitude- bias corresponds to M=40, and largest -in magnitude- bias corresponds to M=5;
(b) Smallest standard deviation corresponds to M=5, and largest standard deviation to M=40.

FIGURE 9. (a) Empirical bias of the bias-corrected estimator $\hat{f}_h(w)$ with M= 10 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$;
(b) Asymptotic standard deviation of $\hat{f}_h(w)$ with M= 10 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$.

FIGURE 10. (a) Empirical bias of the bias-corrected estimator $\hat{f}_h(w)$ with M= 20 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$;
(b) Asymptotic standard deviation of $\hat{f}_h(w)$ with M= 20 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$.

FIGURE 11. (a) Empirical bias of the bias-corrected estimator $\hat{f}_h(w)$ with M=30 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$;
(b) Asymptotic standard deviation of $\hat{f}_h(w)$ with M=30 and $\bar{m}$= 5 as a function of $w \in [0, \pi]$.

FIGURE 12. (a) Empirical bias of the bias-corrected estimator $\hat{f}_h(w)$ with $M=$ 20 and $\bar{m}=$ 10 as a function of $w \in [0, \pi]$;

(b) Asymptotic standard deviation of $\hat{f}_h(w)$ with $M=$ 20 and $\bar{m}=$ 10 as a function of $w \in [0, \pi]$.

FIGURE 13. Mean squared error (MSE) of the Bartlett estimators with $M = 5, 10, 20, 30, 40, 50$ shown as a function of $w \in [0, \pi]$; the MSE curves are nested within one another:

(a) Curves for $M = 5, 10, 20$, with largest MSE corresponding to M=5 and smallest MSE corresponding to M=20;

(b) Curves for $M = 20, 30, 40, 50$, with largest MSE corresponding to M=50 and smallest MSE corresponding to M=20.

FIGURE 14. Mean squared error (MSE) of the bias-corrected estimator $\hat{f}_h(w)$ with $\bar{m}=$ 5 and $M = 10, 20, 30, 40, 50$ shown as a function of $w \in [0, \pi]$; the MSE curves are nested within one another (except in a neighborhood of $w = 0.5$): largest MSE corresponds to M=50 and smallest MSE corresponds to M=10.

FIGURE 15. Mean squared error (MSE) of the bias-corrected estimator $\hat{f}_h(w)$ with $\bar{m}=$ 10 and $M = 20, 30, 40, 50$ shown as a function of $w \in [0, \pi]$; the MSE curves are nested within one another: largest MSE corresponds to M=50 and smallest MSE corresponds to M=20.

FIGURE 16. Mean squared error (MSE) of the bias-corrected estimator $\hat{f}_h(w)$ with $\bar{m}=$ 20 and $M = 30, 40, 50$ shown as a function of $w \in [0, \pi]$; the MSE curves are nested within one another: largest MSE corresponds to M=50 and smallest MSE corresponds to M=30.

FIGURE 17. Mean squared error (MSE) of the remaining bias-corrected estimators $\hat{f}_h(w)$ as a function of $w \in [0, \pi]$; the smallest MSE corresponds to M=40 and $\bar{m}$= 30, the middle line corresponds to M=50 and $\bar{m}$= 30, and the largest MSE corresponds to M=50 and $\bar{m}$=40.

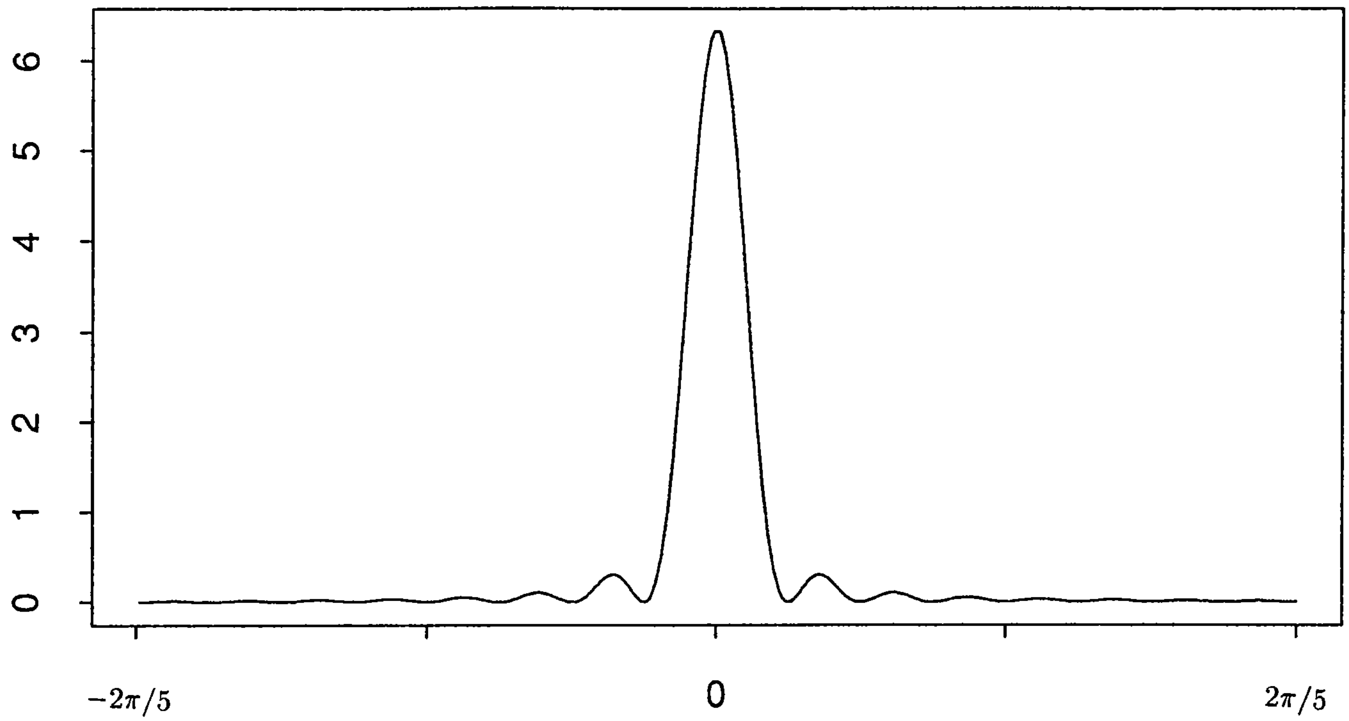


FIGURE 1.

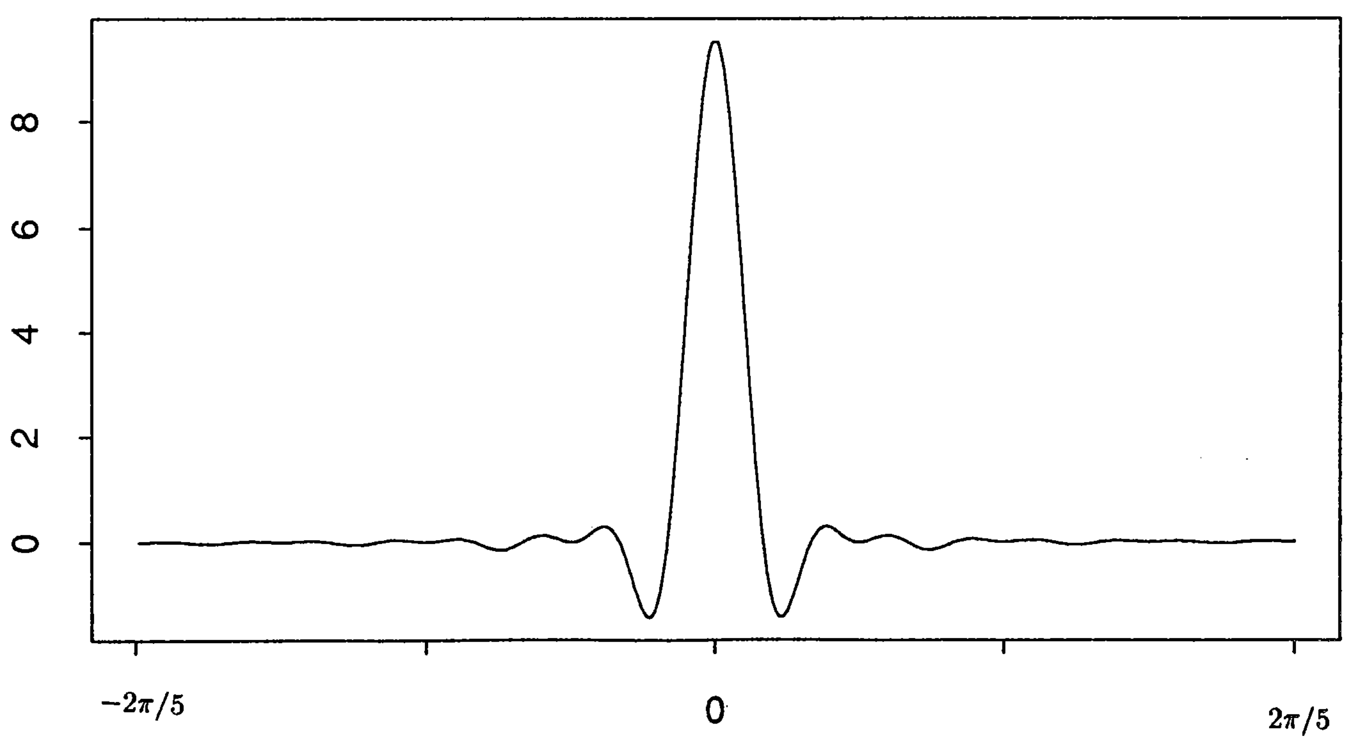


FIGURE 2.

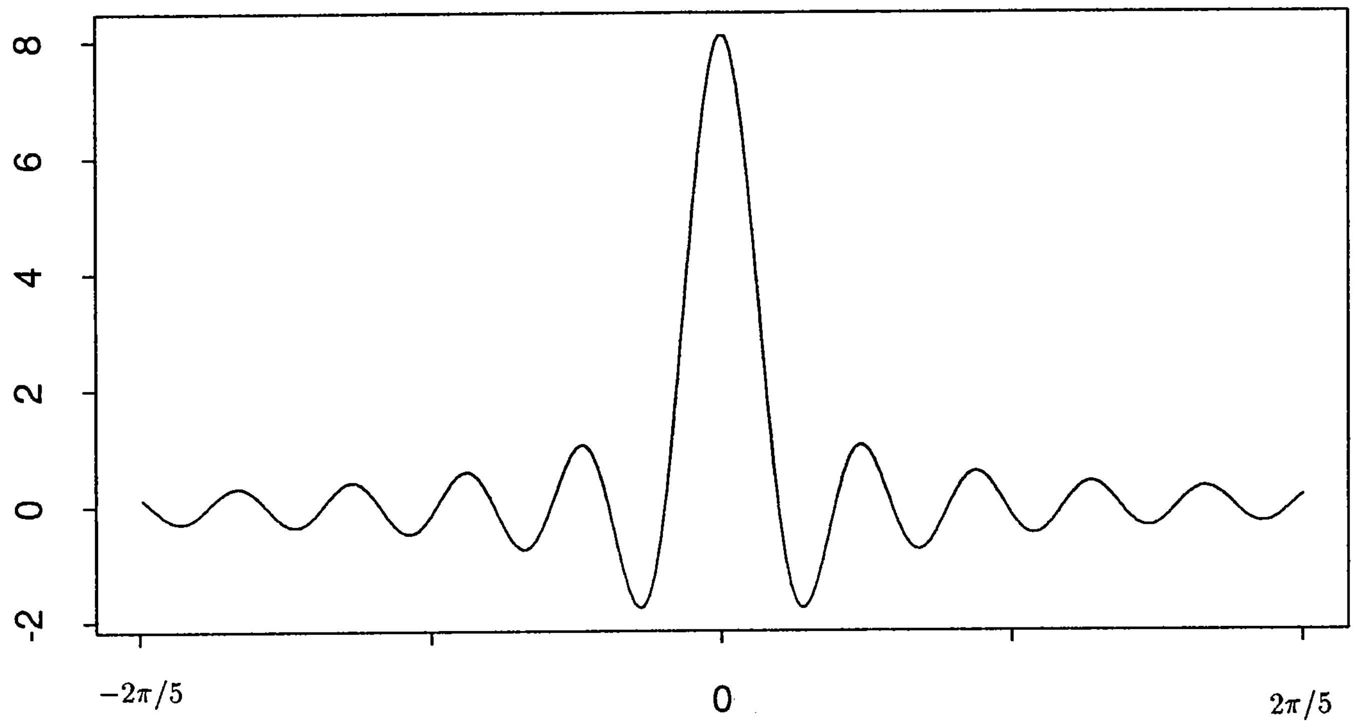
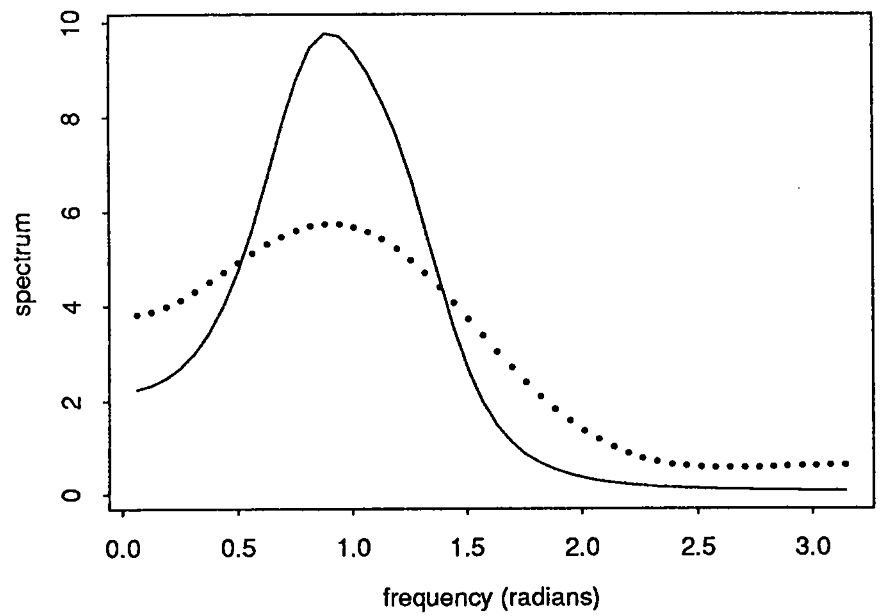
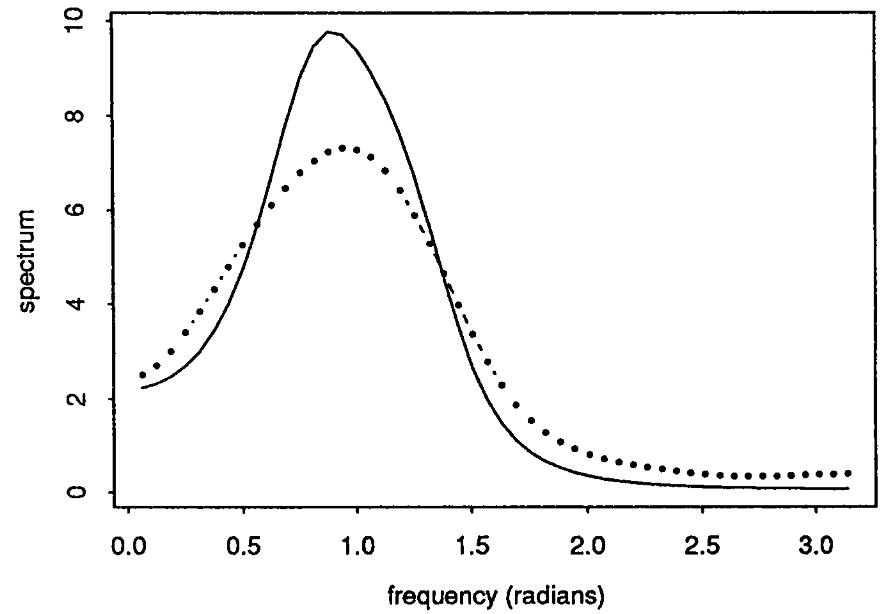


FIGURE 3.

(a)



(b)



(c)

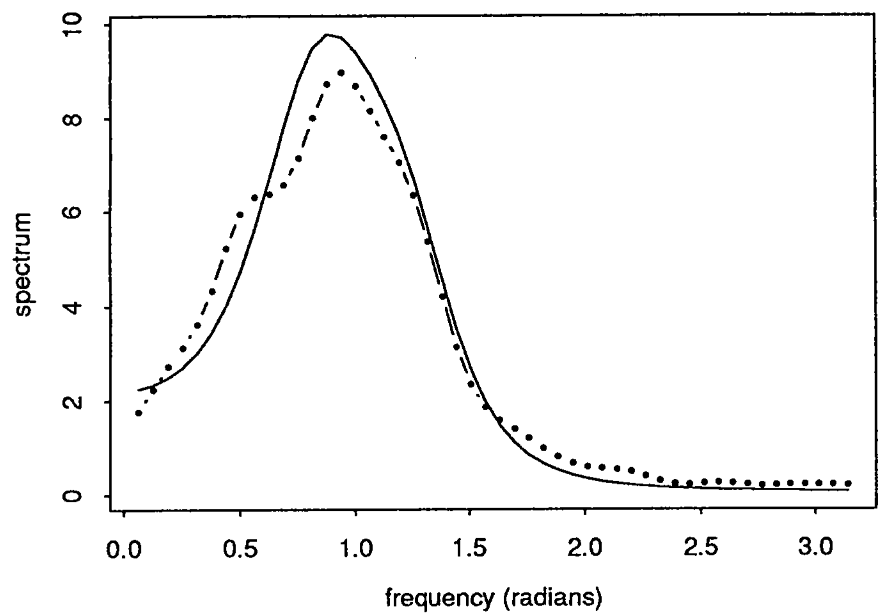
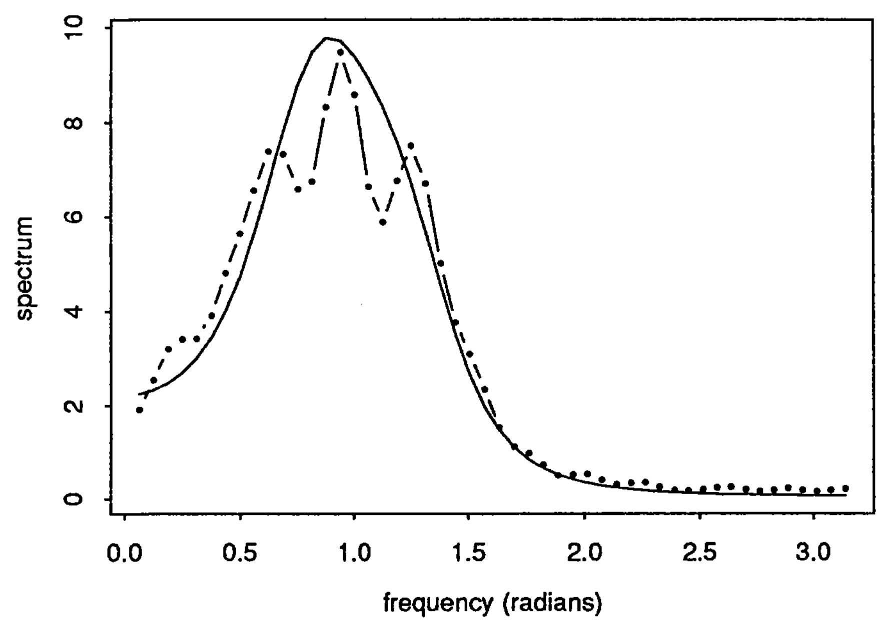


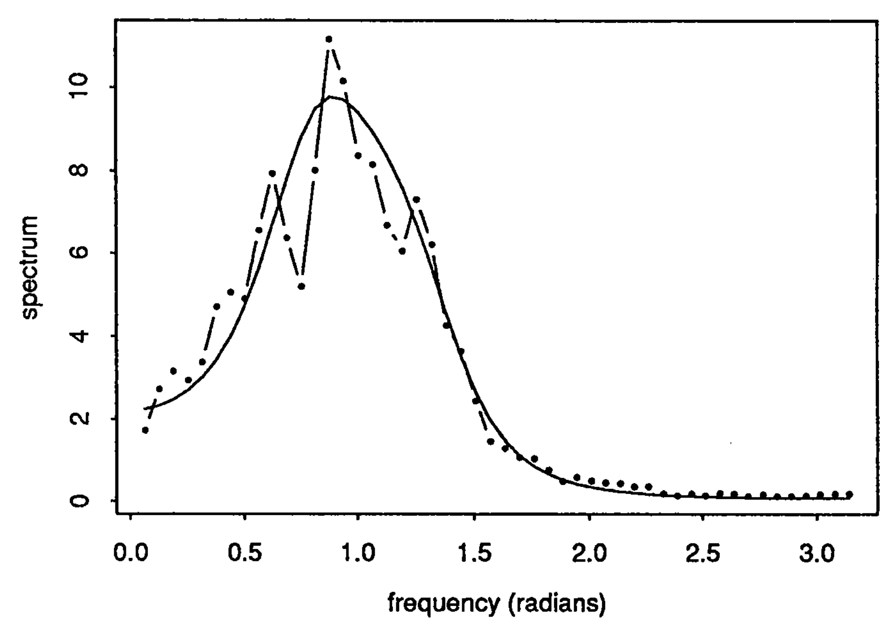
FIGURE 4

(a) (b) (c)

(d)



(e)



(f)

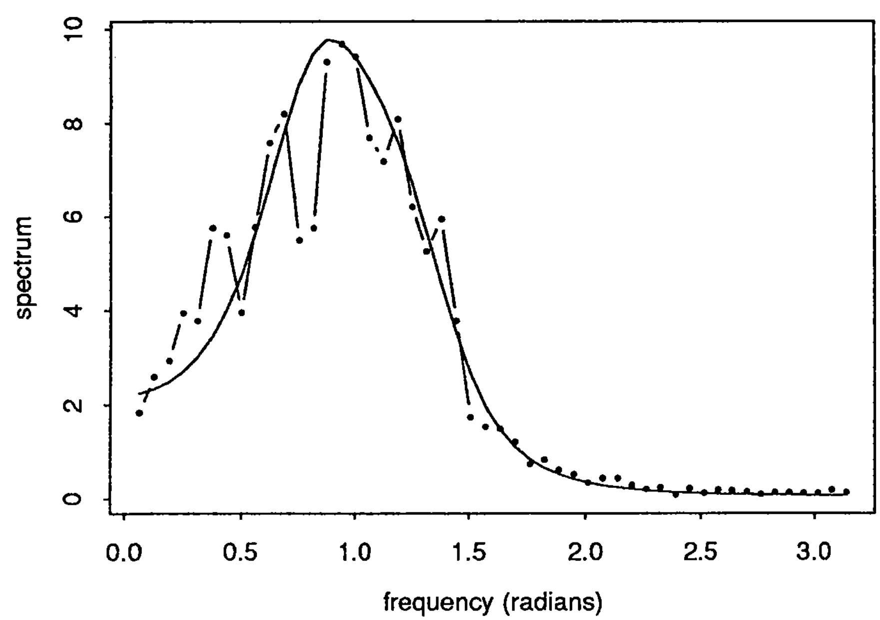


FIGURE 4

(d) (e) (f)

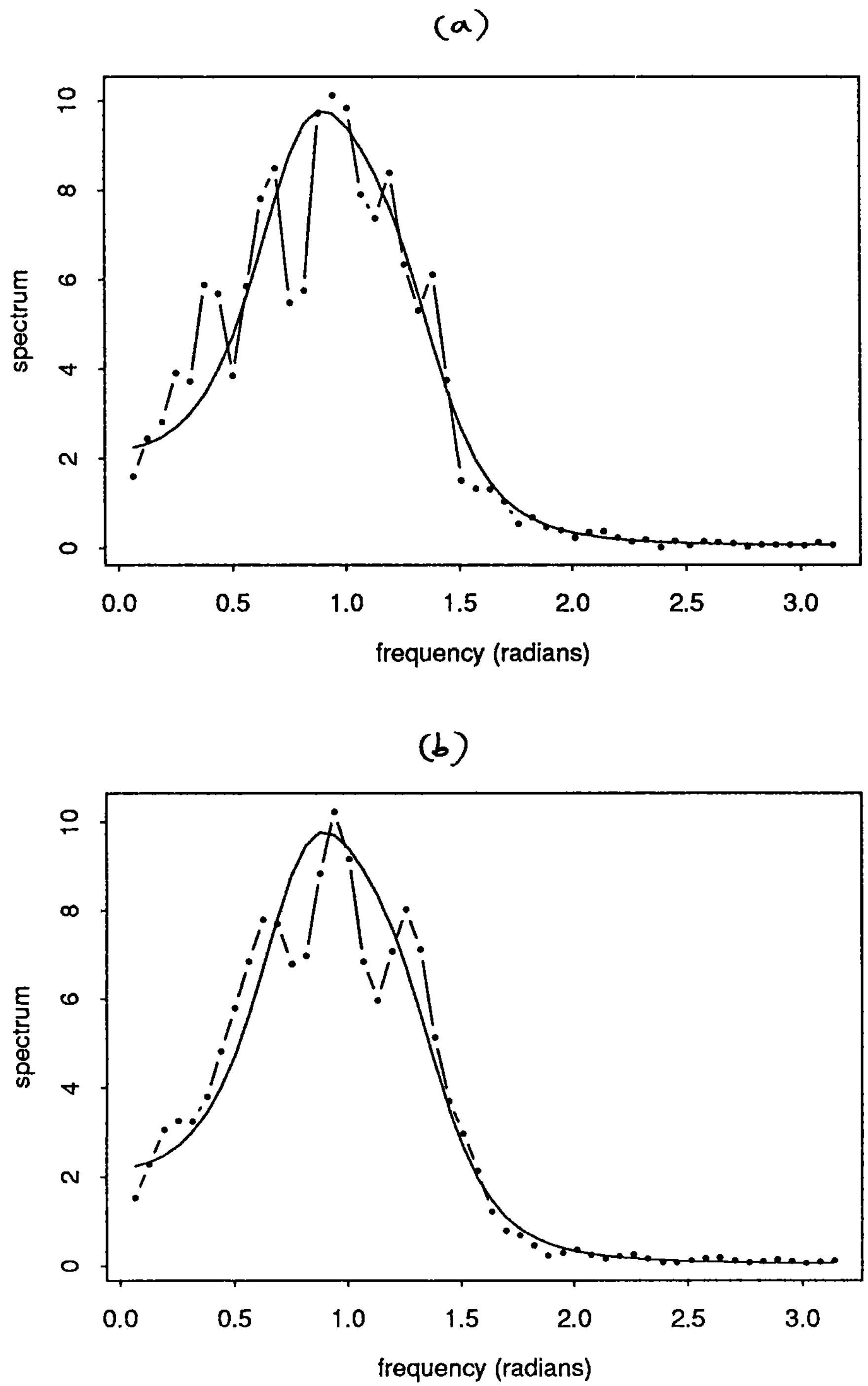

(a)
spectrum
0
2
4
6
8
10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
frequency (radians)
(b)
spectrum
0
2
4
6
8
10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
frequency (radians)

FIGURE 5

(a) (b)

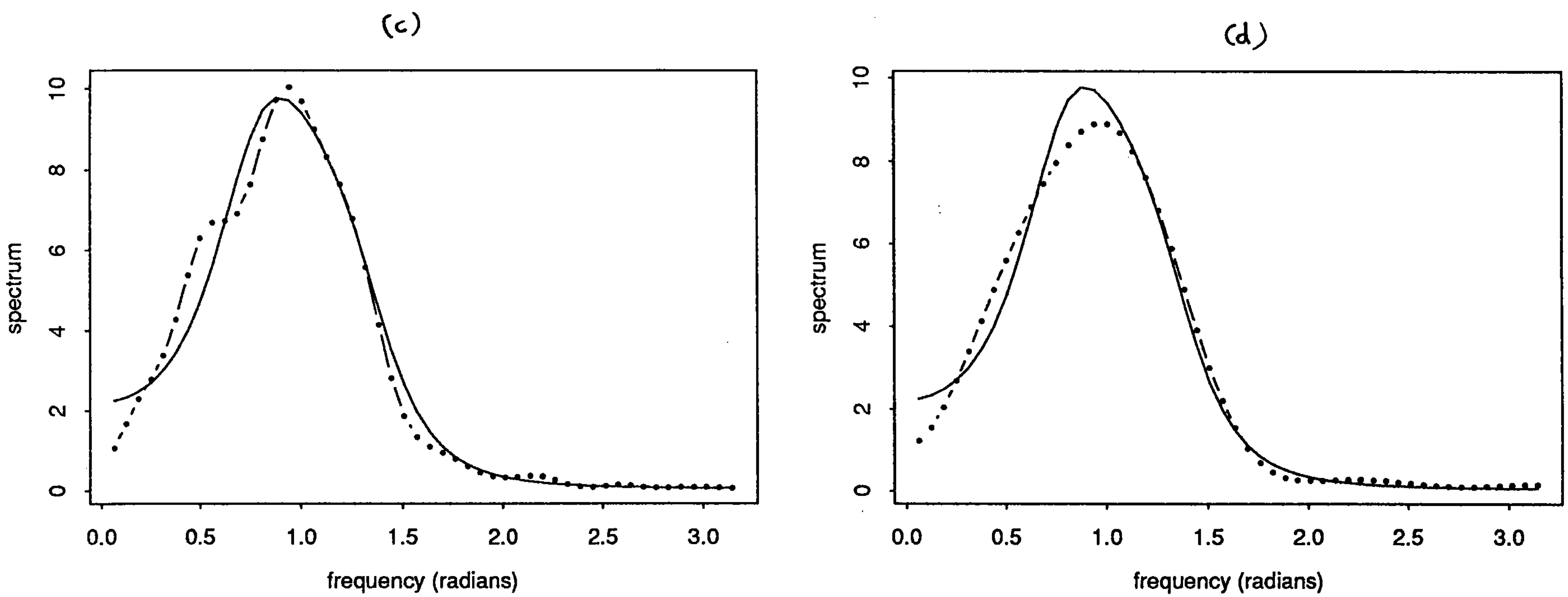
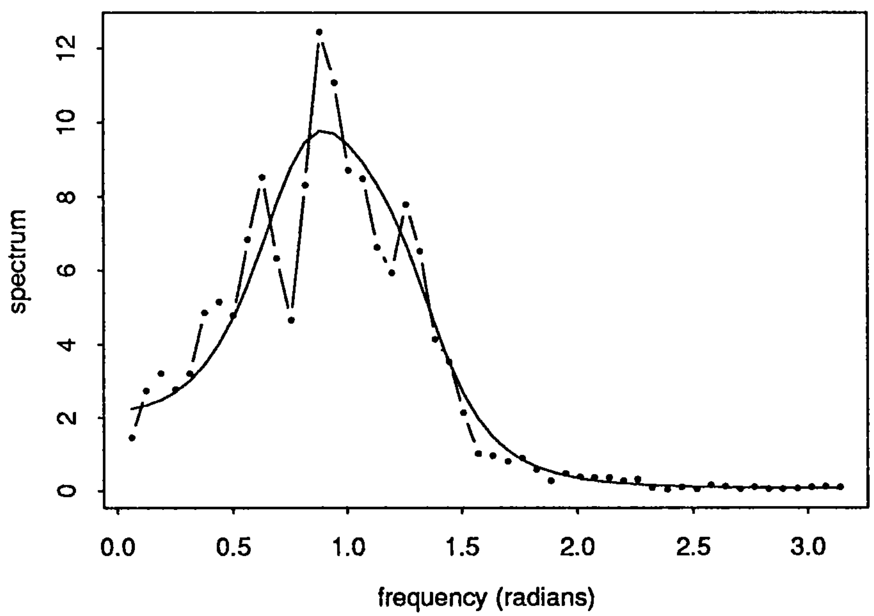


FIGURE 5

(c) (d)

(e)



(f)

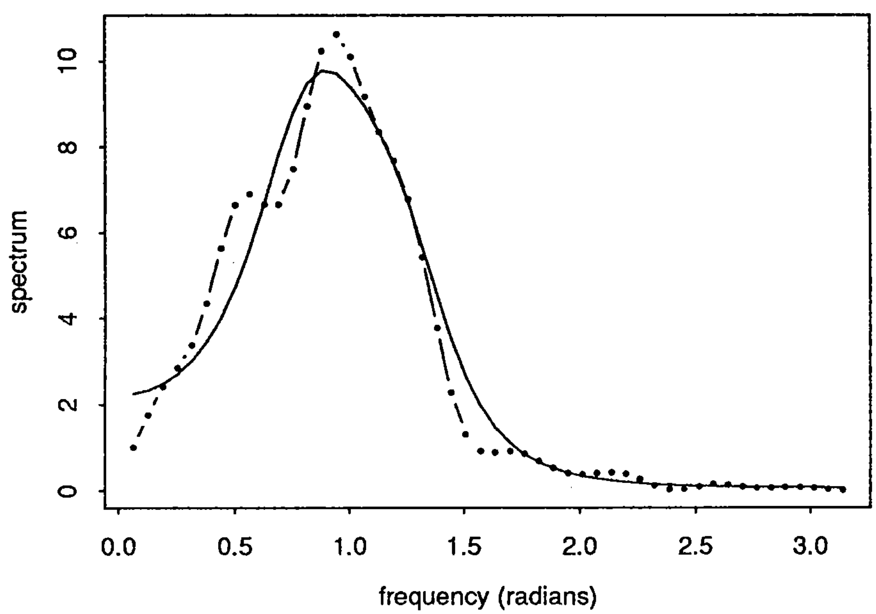


FIGURE 5

(e) (f)

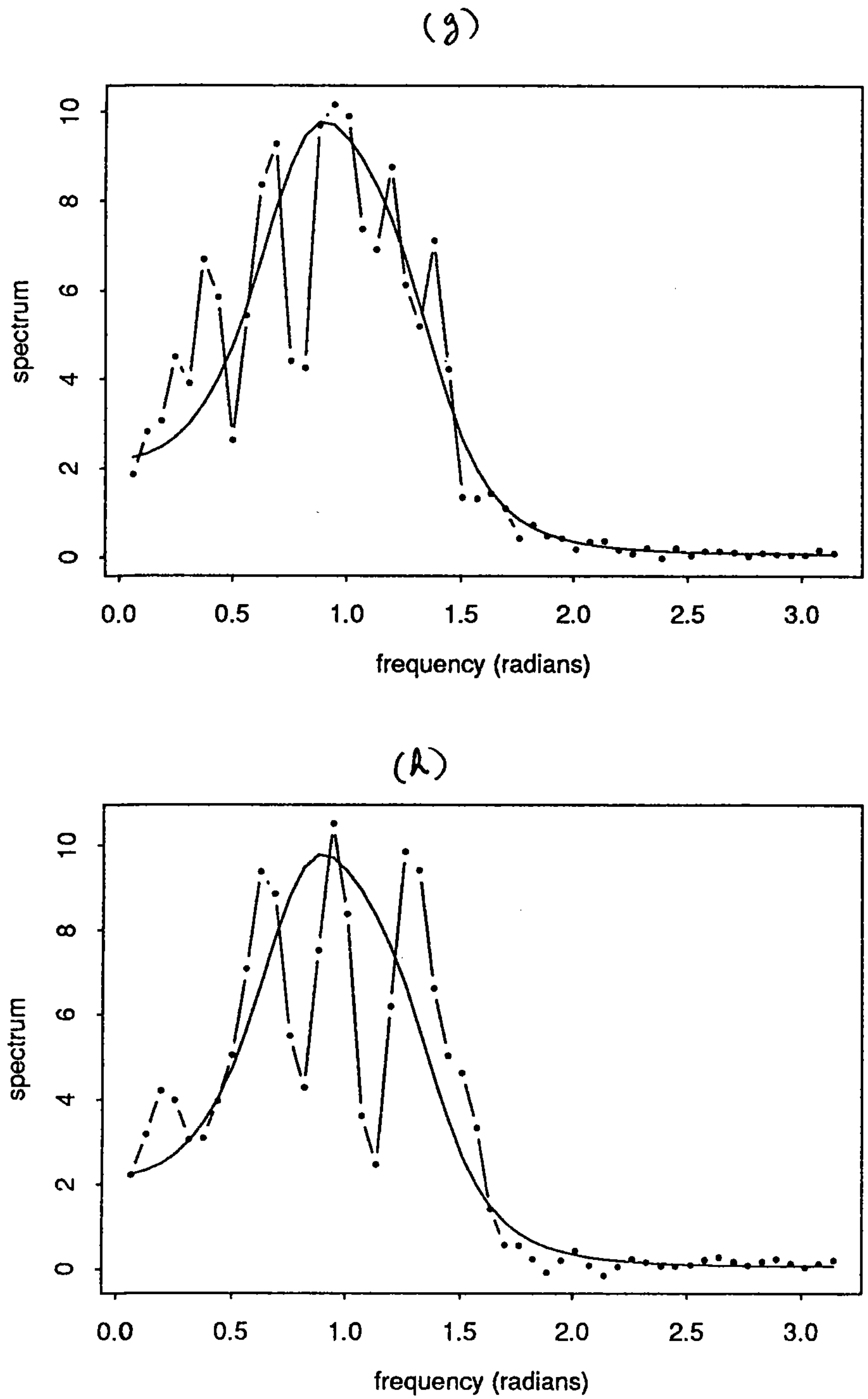
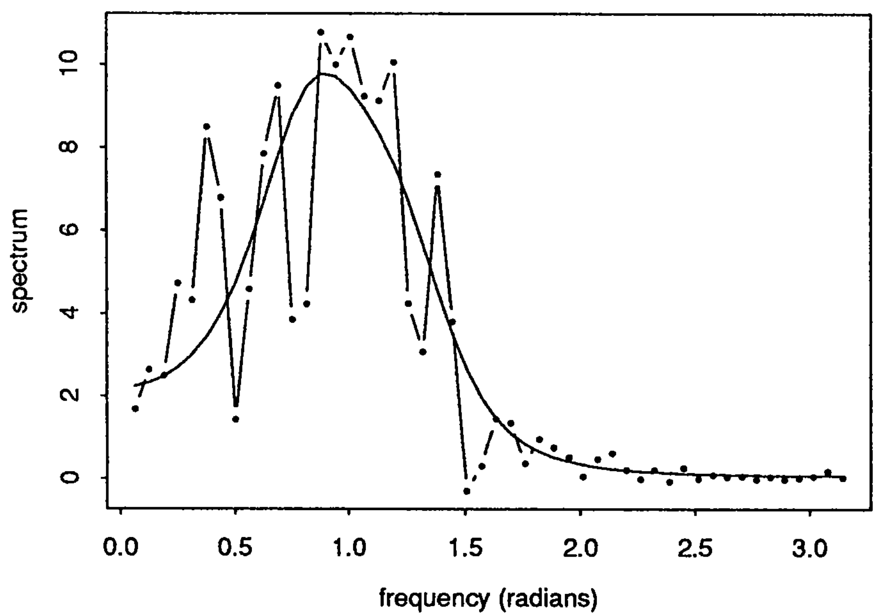

(g)
spectrum
frequency (radians)
(h)
spectrum
frequency (radians)

FIGURE 5

(g) (h)

(i)



(j)

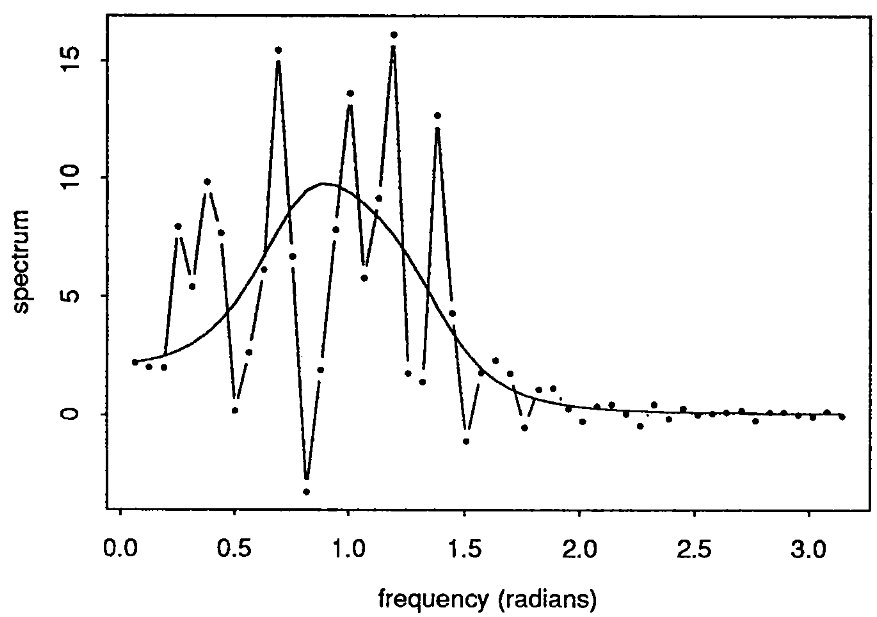


FIGURE 5

(i) (j)

FIGURE 6.

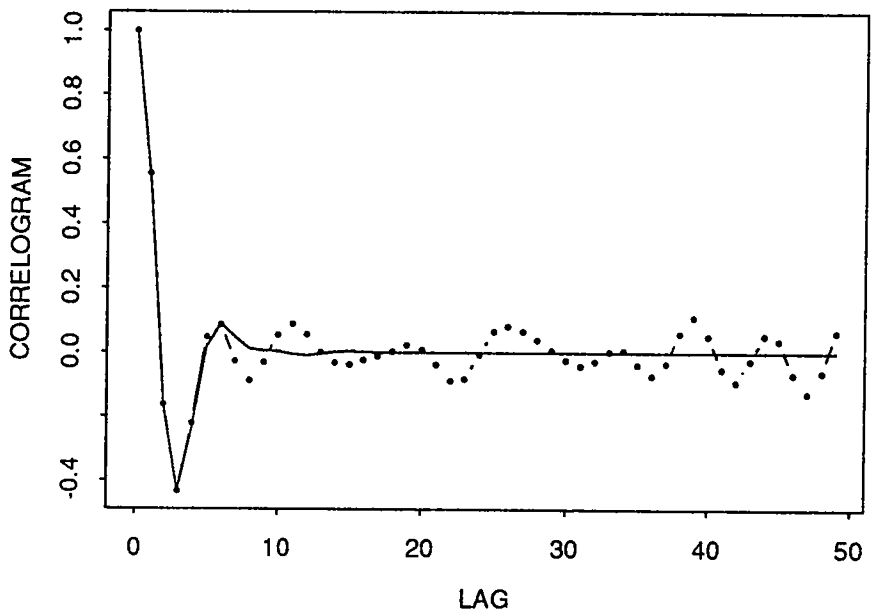


FIGURE 7.

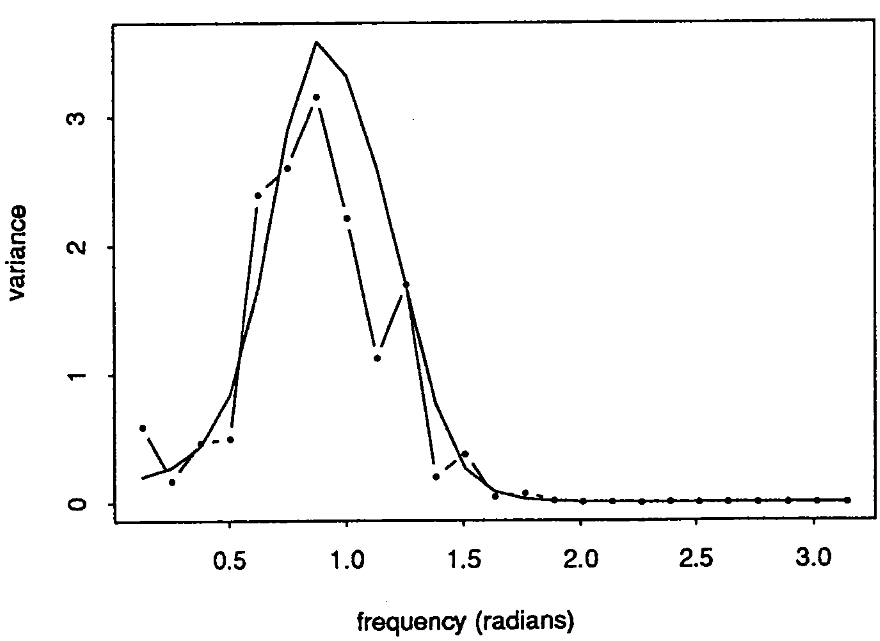


Figure 8.

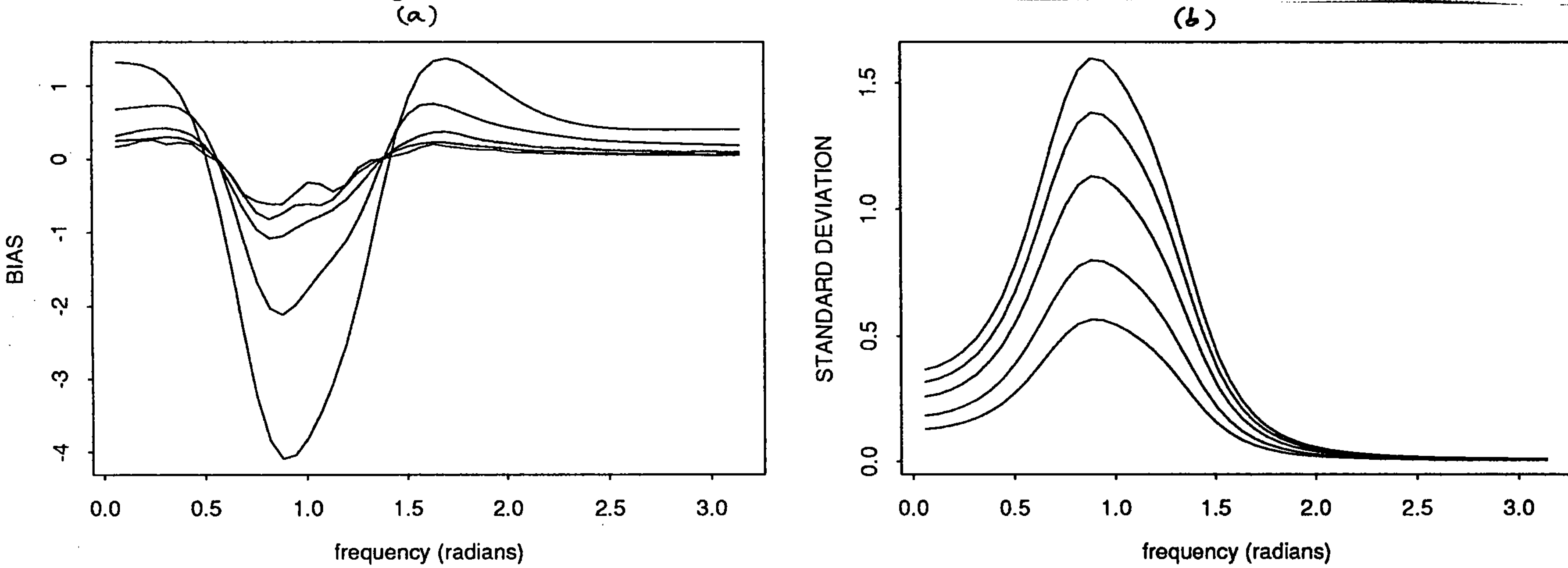

(a)
BIAS
1
0
-1
-2
-3
-4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
frequency (radians)
(b)
STANDARD DEVIATION
0.0
0.5
1.0
1.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
frequency (radians)

FIGURE 9

(a)

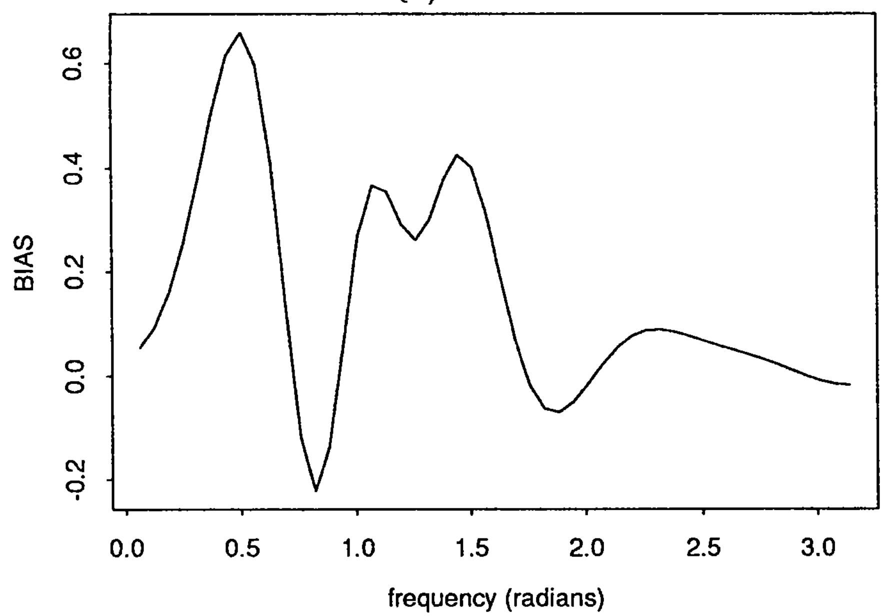


Figure 9

(b)

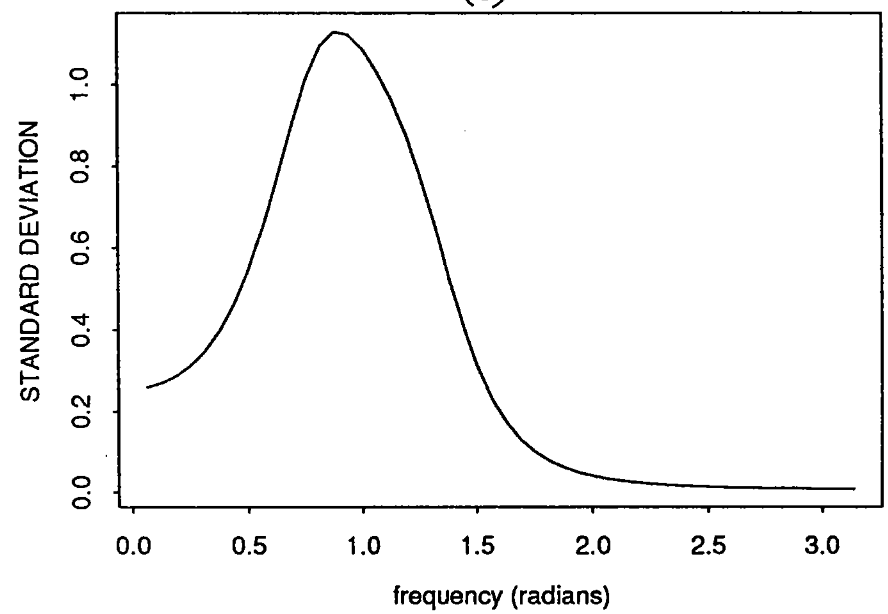


FIGURE 10

(a)

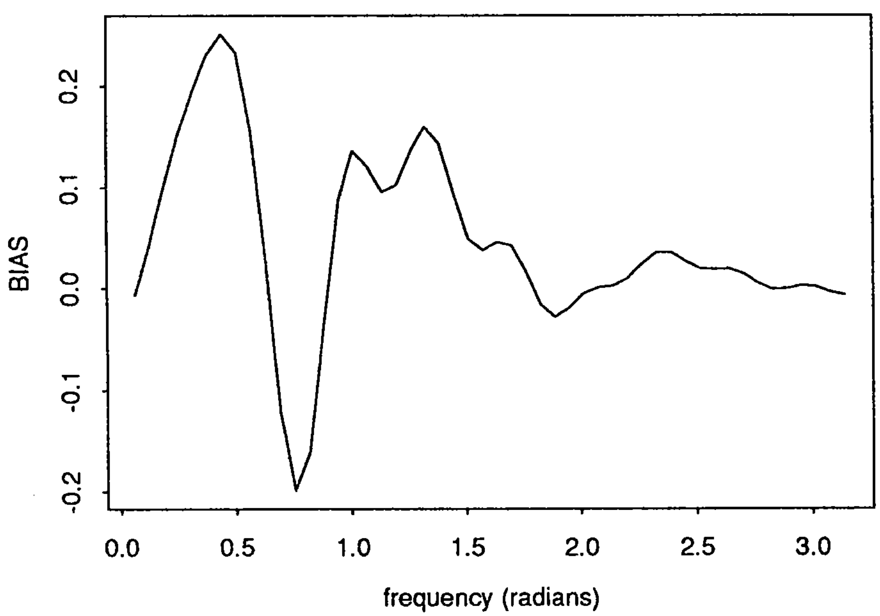


Figure 10

(b)

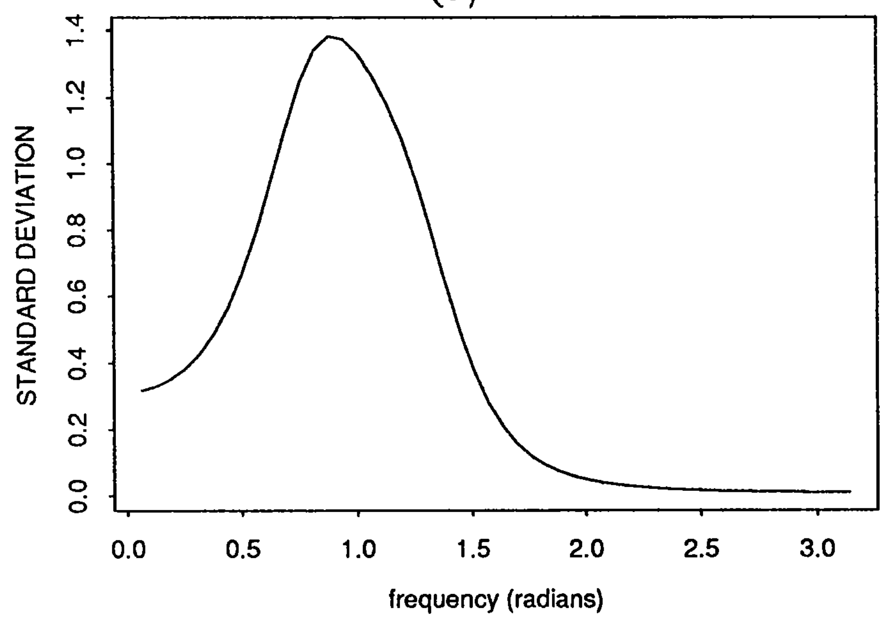


FIGURE 11 (a)

Figure 11 (b)

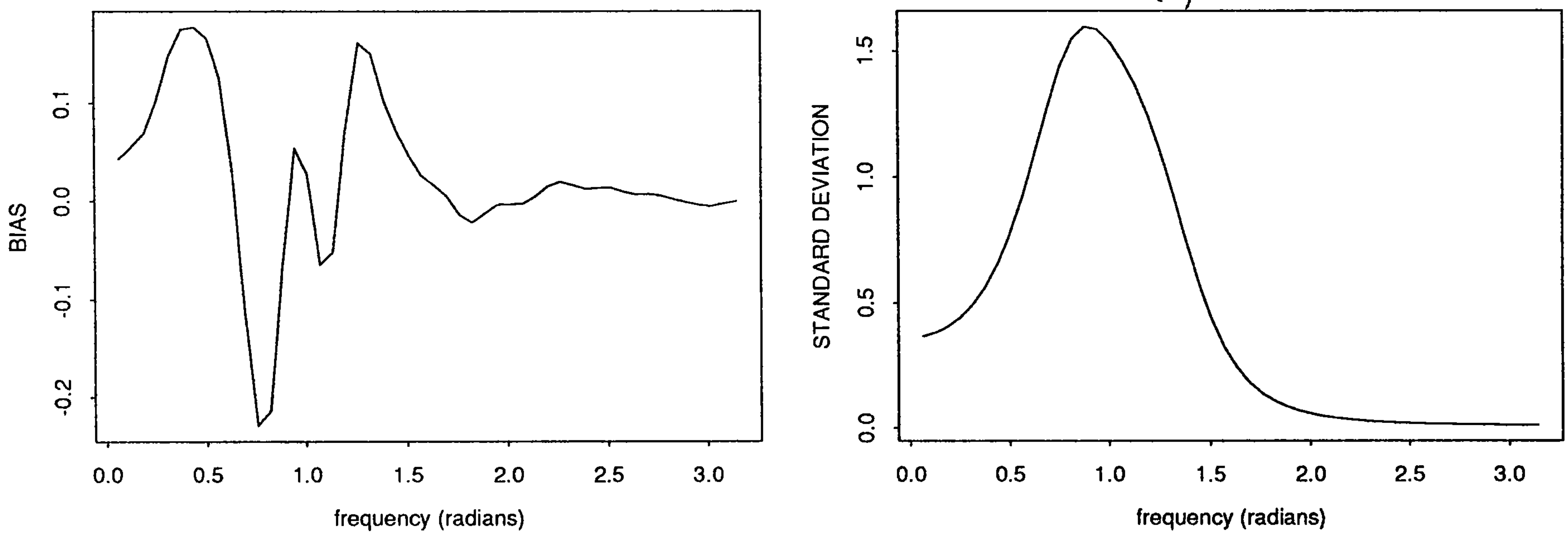


FIGURE 12 (a)

Figure 12: (b)

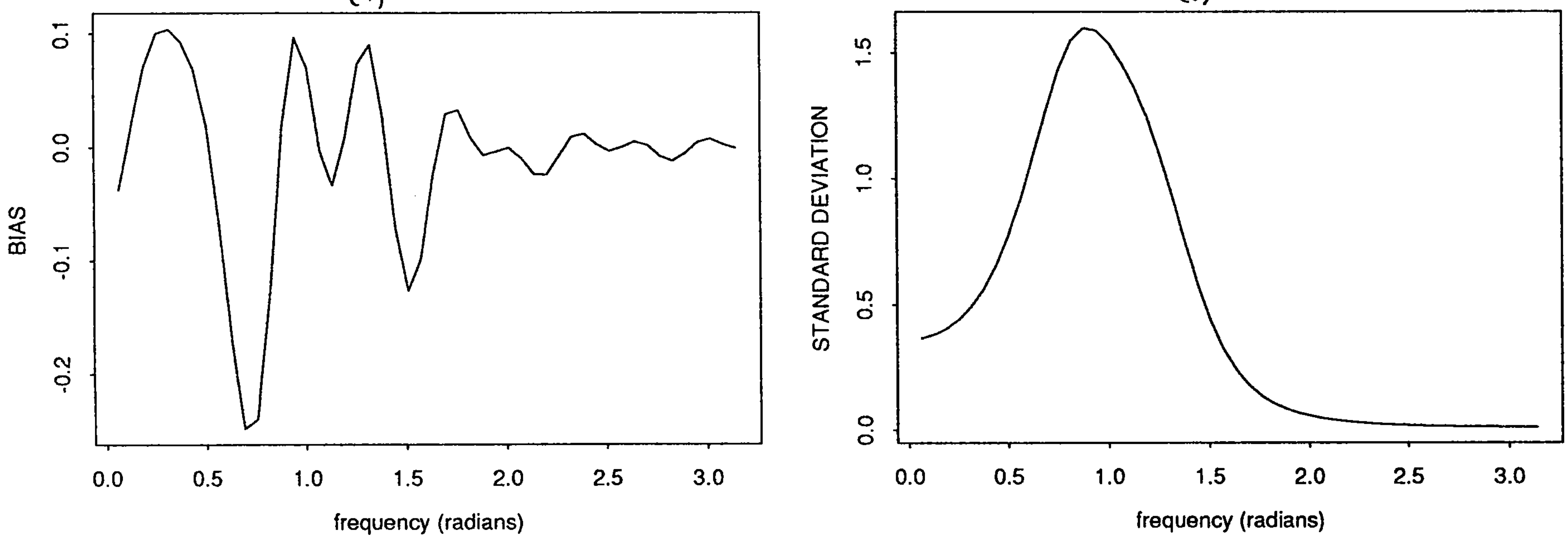


FIGURE 13 (a)

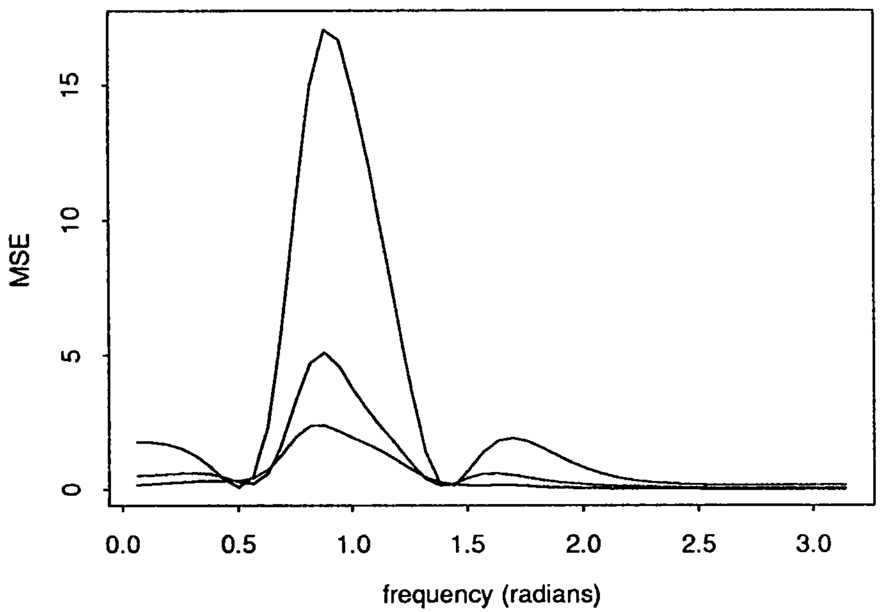


figure 13 (b)

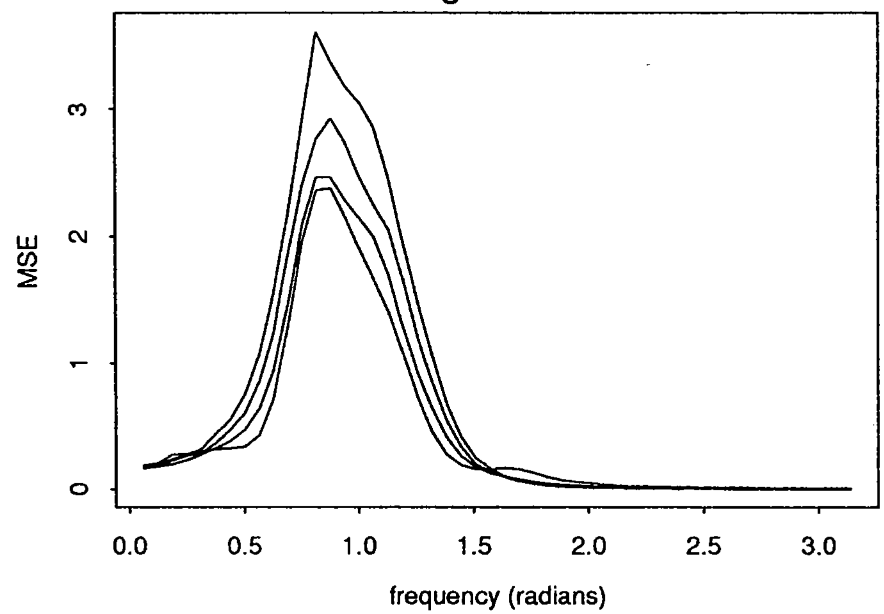


figure 14

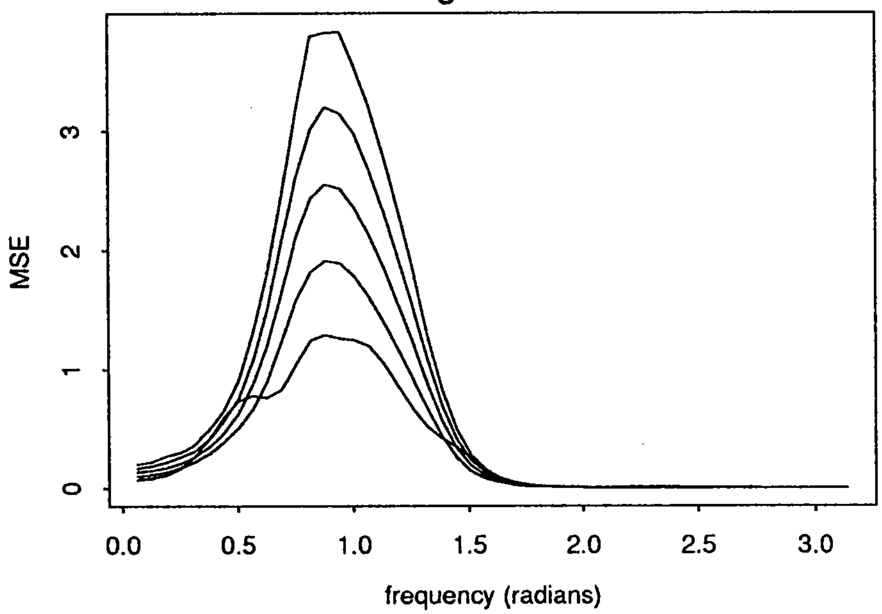


figure 15

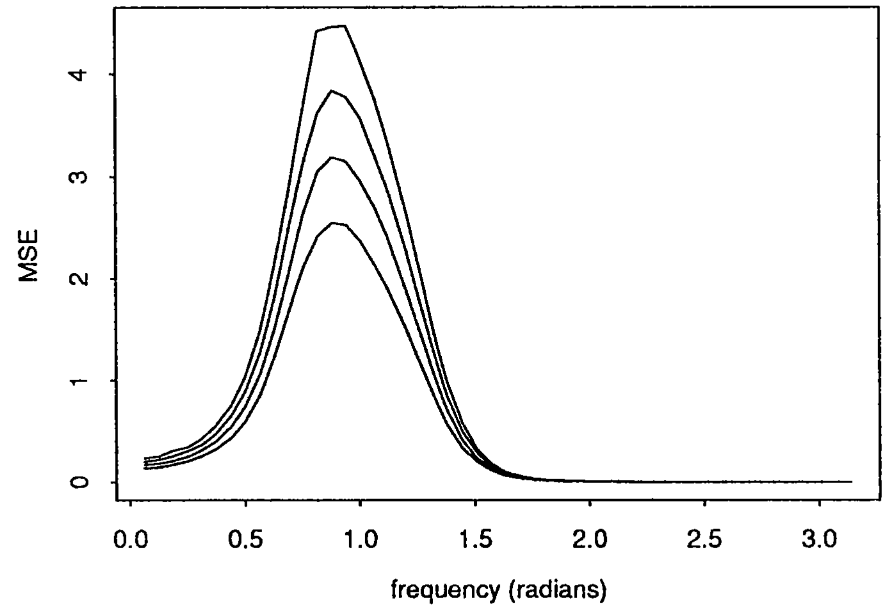


figure 16

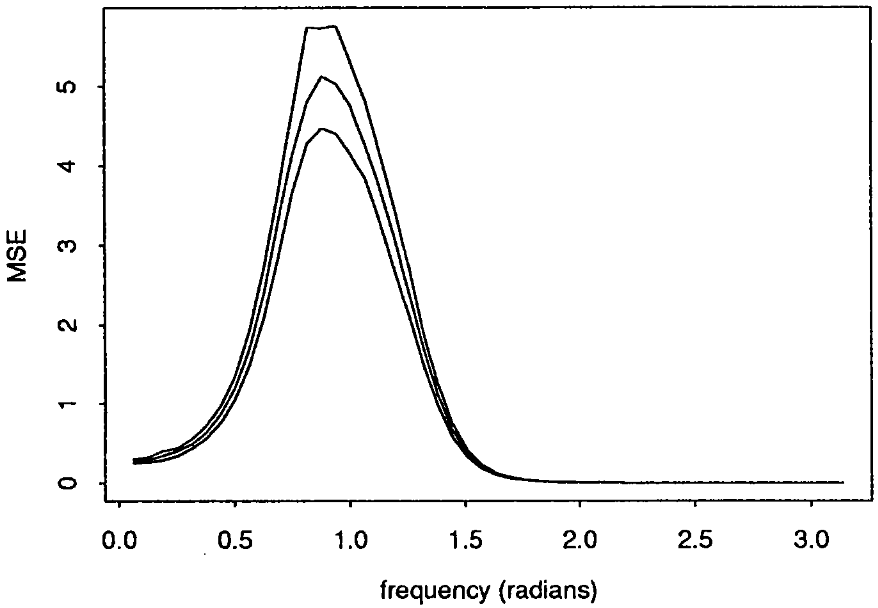


figure 17