BIAS-CORRECTED NONPARAMETRIC SPECTRAL ESTIMATION

by

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

Technical Report #92-50

Department of Statistics
Purdue University

November 1992

Bias-corrected nonparametric spectral estimation

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 extensively studied in recent years. However, the most popular spectral estimators, e.g., 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 of the bias of a nonparametric spectral estimator is presented. This will result in a general proposal for bias corrected estimates that are related to kernel type estimators with lag-windows of trapezoidal shape. An important application of the proposed bias reduction scheme is found in improving the standard error estimates obtained by resampling and subsampling time series data.

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 m-dependent data. The finite-sample performance of the trapezoidal lag-window estimator will be assessed by simulation in a follow-up paper.

Keywords. Bartlett's window, bias reduction, bootstrap, jackknife, 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., $EX_t = 0$. Suppose also 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[1] (and most intuitive) proposals for consistent estimation of $f(w)$ at some point $w \in [-\pi, \pi]$ was given by Bartlett (1946) and goes as follows. Let

$$T_{i,M,L}(w) = \frac{1}{2\pi M}\left|\sum_{t=L(i-1)+1}^{L(i-1)+M} X_t e^{-jtw}\right|^2,$$

that is, $T_{i,M,L}(w)$ is the periodogram of block $\{X_{L(i-1)+1}, \ldots, X_{L(i-1)+M}\}$ of the data, where L, M are integers depending on the sample size N; M is the block's size, and L is the amount of 'lag' between the starting points of block i and block $i+1$. Now define

$$\bar{T}_N(w) = \frac{1}{Q}\sum_{i=1}^{Q} T_{i,M,L}(w), \tag{1}$$

where $Q = [\frac{N-M}{L}] + 1$, and $[\cdot]$ is the integer part.

Under regularity conditions (cf. Priestley(1981), Zhurbenko (1986)), $\bar{T}_N(w)$ 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, e.g., conditions on the smoothness of $f(w)$, and conditions on the design parameters, typically $M \to \infty$, but with $M/N \to 0$ and $L/M \to 0$ as $N \to \infty$.

Notably, there is an alternative (and asymptotically equivalent) way to compute $\bar{T}_N(w)$, which amounts to a kernel smoothed spectral estimator using the triangular (Bartlett's) kernel, i.e.,

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

where $\hat{R}(s) = \frac{1}{N}\sum_{k=1}^{N-s} X_kX_{k+s}$ is the usual sample autocovariance. It can be shown (cf.

[1] A proposal from a different prespective was also given by Daniell (1946).

Priestley (1981)) that $\tilde{f}(w) \simeq \bar{T}_N(w)$, and that the two estimators have similar asymptotic properties (bias, variance, etc.).

In particular, the asymptotic variance of either $\tilde{f}(w)$ or $\bar{T}_N(w)$ is approximately $\frac{2}{3}\frac{M}{N}f^2(w)(1+\eta(w))$, where $\eta(w) = 0$ if $w \neq 0(mod\pi)$ and $\eta(w) = 1$ if $w = 0, \pm\pi$. In addition, the large sample distribution of either $\sqrt{N/M}(\tilde{f}(w) - f(w))$ or of $\sqrt{N/M}(\bar{T}_N(w) - f(w))$ is the normal $N(0, \frac{2}{3}f^2(w)(1 + \eta(w)))$, provided $M \sim aN^\beta$, for some $a > 0$ and $0 > \beta > 1/3$, as $N \to \infty$; following the usual convention, the notation $M \sim aN^\beta$ is a short-hand for $M/N^\beta \to a$, as $N \to \infty$.

However, the asymptotic bias of $\tilde{f}(w)$ or $\bar{T}_N(w)$ is rather large[2] (of approximate order c_1/M, for some $c_1 > 0$ depending on w and on f), and subsequent research efforts were expended to obtain spectral estimators with smaller bias. These efforts were pointed to two main directions that are direct analogs of the construction of $\tilde{f}(w)$ and $\bar{T}_N(w)$, namely: (a) using a different kernel, one that is smoother near the origin, than the triangular one for smoothing (cf. Grenander and Rosenblatt(1957), Blackman and Tukey(1959), Parzen(1957a, 1957b), Priestley(1981)), and (b) averaging the periodograms of consecutive blocks of data that are 'tapered' by an appropriate smooth function that starts and ends on zero (cf. Welch(1967), Brillinger(1975), Zhurbenko (1986)).

As in the case of Bartlett's kernel, these two directions are analogous, and there is a correspondence between a kernel for method (a) and a tapering function for method (b) that makes the estimators obtained by the two methods equivalent (cf. Priestley(1981)). By appropriately choosing the kernel (or the tapering function), 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 perpective 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 general 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 $O(1/M^r)$, where r can be intuitively interpreted as the number

[2]More precisely, $E\tilde{f}(w) - f(w) = O(1/M)$, and $E\bar{T}_N(w) - f(w) = O(\log M/M)$, under the assumption that f is continuously differentiable.

of derivatives that f has.

Attention will be focused to Bartlett's estimator because it is both the simplest 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,b)), and in the steady state simulation literature (cf. Meketon and Schmeiser (1984), Welch (1987), Song and Schmeiser (1988, 1992)). However, with obvious modifications, the same intuitive procedure can be carried out for bias reduction of other spectral estimators, and indeed even for bias reduction of kernel smoothed probability density estimators in a different context.

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. This basic idea is generalized in Section 3; see Sections 3.1 and 3.2. Section 3.3 is concerned with the positivity of the proposed estimator, and Section 3.4 establishes its $\sqrt{N}$-consistency in the case of Moving Average data. Finally, Section 3.5 addresses the most important problem of choosing the bandwidth of the spectral estimator in practice, and Section 4 contains some comments and conclusions. Proofs of all results are placed in the Appendix.

2. Bias reduction for Bartlett's spectral estimator: the $2f - f$ formula

2.1. Some heuristic ideas. From the data $X_1, \ldots, X_N$ construct the Bartlett spectral estimator $\tilde{f}(w)$ at point w as given in equation (2) 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 $\tilde{f}(w)$ to be approximately $\widehat{Bias}(\tilde{f}(w)) = \bar{f}(w) - \tilde{f}(w)$. 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 (see Figure 1). 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 estimate of the bias of $\tilde{f}(w)$ will at least have the right sign. The magnitude of this estimate of bias is of course also of great importance, since one might 'over-correct' by overestimating the (absolute value of the) bias. This is obviously related to how one chooses $\bar{m}$, i.e., how 'over-smoothed' $\bar{f}(w)$ is.

2.2. Best choice of $\bar{m}$ in the $2f - f$ formula. It is worth noting that the estimator $\hat{f}(w)$ as defined in equation (4) is actually a nonparametric spectral estimator of the lag-window type, i.e., $\hat{f}(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \hat{\lambda}(\frac{s}{M})\hat{R}(s)e^{-jsw}$, where

$$\hat{\lambda}(\frac{s}{M}) = \begin{cases} 2(1 - \frac{|s|}{M}) - (1 - \frac{|s|}{\bar{m}}) & \text{for } |s| \leq \bar{m} \\ 2(1 - \frac{|s|}{M}) & \text{for } \bar{m} < |s| \leq M \\ 0 & \text{for } |s| > M. \end{cases}$$

In Figure 2, the lag-window $\hat{\lambda}(\frac{s}{M})$ is pictured for different values of the ratio $\bar{m}/M$. It is apparent that taking $\bar{m} << M/2$ would not be recommended, as the lag-window would then give more weight to $\hat{R}(s)$ for s around $\bar{m}$, than to $\hat{R}(s)$ for s around zero. Intuitively, taking $\bar{m} << M/2$ would lead to 'over-correcting' the bias, by overestimating it. Taking $\bar{m} >> M/2$ actually leads to $\hat{f}(w)$ being very close to the Bartlett estimate $\tilde{f}(w)$, i.e., there is not 'enough' bias correction.

It turns out that the best choice of $\bar{m}$ is $M/2$ or $(M+1)/2$, according to whether M is even or odd respectively. This choice makes the lag-window $\hat{\lambda}(\frac{s}{M})$ to be of trapezoidal shape (see Figure 2). Since one can always take M to be even, we will henceforth assume so to make the discussion concrete. As a matter of course, the asymptotic considerations are all the same if M is even or odd, and if $\bar{m}$ is $M/2$ or $(M+1)/2$.

To justify that the choice of $\bar{m} = M/2$ achieves the claimed bias reduction, the following theorem is presented. Note that the choice of $\bar{m}$ here is based only on bias considerations, since the large-sample variance of $\hat{f}(w)$ is of order $O(M/N)$ whatever the choice of $\bar{m}/M$ may be.

Theorem 1 *Let w be any point in $[-\pi, \pi]$, and assume*[3] *that $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$, for some positive integer r. Also assume that $\bar{m} \sim M/2$, and that $M \to \infty$, as $N \to \infty$, but with $M^r/N \to 0$. Then*

$$Bias(\hat{f}(w)) = O(1/M^r). \tag{5}$$

Suppose in addition that the time series $\{X_t\}$ is such[4] *that the Bartlett spectral estimator $\tilde{f}(w)$ has a large-sample variance of order $O(M/N)$. Then we also have*

$$Var(\hat{f}(w)) = O(M/N). \tag{6}$$

The interpretation of Theorem 1 is that for the case $r > 1$, the bias of $\hat{f}(w)$ is smaller than the bias of the Bartlett estimator $\tilde{f}(w)$ by orders of magnitude, and thus the '$2f - f$' formula of equation (4) achieves its purpose of bias-correction without changing the asymptotic order of the variance. If $r = 1$, i.e., if f is not smooth enough, then the performance of $\hat{f}(w)$ is similar to that of the Bartlett estimator $\tilde{f}(w)$.

[3] Note that the assumption $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$ implies that f has r continuous derivatives; conversely, if f has $r+1$ derivatives and the $(r+1)$th derivative is square-integrable, then $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$ follows (cf. Katznelson (1968)).

[4] There is a variety of assumptions to guarantee that $Var(\tilde{f}(w)) = O(M/N)$; for example (cf. Priestley (1981)) a sufficient condition is that $\{X_t\}$ is a linear process given by $X_t = \sum_{i=-\infty}^{\infty} \theta_i Z_{t-i}$, 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 $\sum_{i=-\infty}^{\infty} |i|^{1/2} |\theta_i| < \infty$.

3. Bias reduction for Bartlett's spectral estimator: a more general formula

3.1. The general bias correction formula. As it was argued in Section 2.1, it is intuitively clear that the 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 new bias-corrected estimator given by

$$\check{f}(w) \equiv \tilde{f}(w) - \widehat{\widehat{Bias}}(\tilde{f}(w)) = \tilde{f}(w) - h\widehat{Bias}(\tilde{f}(w)) = (h+1)\tilde{f}(w) - h\bar{f}(w). \tag{7}$$

This new estimator $\check{f}(w)$ is also of the lag-window type, i.e., $\check{f}(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \check{\lambda}(\frac{s}{M})\hat{R}(s)e^{-jsw}$, where the lag-window $\check{\lambda}(\frac{s}{M})$ is defined by

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

By similar considerations as in Section 2.2, it follows that the best choice of $\bar{m}$ in this case is $\bar{m}/M \sim h/(h+1)$, in which case the lag-window $\check{\lambda}(\frac{s}{M})$ is flat around the origin, and possesses the general trapezoidal shape (see Figure 3). Hence, given $\bar{m}$ and M, the value of h is also determined and is $h = \bar{m}/(M - \bar{m})$.

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 (7) as the linear combination of spectral estimators with minimum mean squared error. This most desirable feature of the proposed estimator will be investigated in the next section.

3.2. Performance of the general trapezoidal lag-window. It is apparent that the proposed bias-corrected estimators $\hat{f}(w)$ (with $\bar{m} \sim M/2$) and $\check{f}(w)$ (with $\bar{m} \sim Mh/(h+1)$) are special cases of the nonparametric spectral estimator

$$\hat{f}_h(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \lambda(\frac{s}{M})\hat{R}(s)e^{-jsw}, \tag{8}$$

with the general trapezoidal lag-window $\lambda(\frac{s}{M})$ defined by

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

A more precise notation for the above trapezoidal lag-window might be $\lambda_h(\frac{s}{M})$, where $h = \bar{m}/(M - \bar{m})$, but we will generally drop the subscript h since it does not lead to confusion. As elaborated upon in Section 3.1, 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

$$\hat{f}_h(w) = (h+1)\tilde{f}(w) - h\bar{f}(w). \tag{9}$$

The assumption that h is a constant is equivalent to assuming that $\bar{m}$ is proportional to M. If $h = 1$, then the estimator $\hat{f}_h(w)$ is identical to $\hat{f}(w)$ defined in equation (4). At the extreme points where $h = 0$ or $h = \infty$, that correspond to $\bar{m} = 0$ and $\bar{m} = M$ respectively, $\hat{f}_h(w)$ reduces to a Bartlett estimator or the 'truncated periodogram' $\frac{1}{2\pi}\sum_{s=-M}^{M}\hat{R}(s)e^{-jsw}$ respectively.

Note (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

$$\Lambda_h(w) \equiv \frac{1}{2\pi}\sum_{s=-M}^{M}\lambda(\frac{s}{M})e^{-jsw} \tag{10}$$

is the so-called spectral window corresponding to the lag-window $\lambda(\frac{s}{M})$. 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,

$$\begin{aligned}\Lambda_h(w) &= \frac{h+1}{2\pi M}\left(\frac{\sin(Mw/2)}{\sin(w/2)}\right)^2 - \frac{h}{2\pi\bar{m}}\left(\frac{\sin(\bar{m}w/2)}{\sin(w/2)}\right)^2 \\ &= \frac{1}{2\pi(M-\bar{m})}\left\{\left(\frac{\sin(Mw/2)}{\sin(w/2)}\right)^2 - \left(\frac{\sin(\bar{m}w/2)}{\sin(w/2)}\right)^2\right\}.\end{aligned} \tag{11}$$

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 (11), the spectral kernel $\Lambda_h(w)$, for $0 \leq h < \infty$ 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.

The fact that $\int_{-\pi}^{\pi} w^k \Lambda_h(w) dw$ is of smaller order of magnitude than $\int_{-\pi}^{\pi} w^k \Lambda_0(w) dw$ seems to indicate that $\hat{f}_h(w)$ has smaller bias than $\hat{f}_0(w) = \tilde{f}(w)$. This is indeed true, and the following theorem quantifies the bias-variance performance of $\hat{f}_h(w)$ as an estimator of $f(w)$, for $0 < h < \infty$. If f is smooth enough, the estimator $\hat{f}_h(w)$ is shown to have smaller (by orders of magnitude) bias than the Bartlett estimator $\tilde{f}(w)$, while its variance remains of the same order of magnitude. From the theorem's proof it is obvious that the small bias of $\hat{f}_h(w)$ is a consequence of the 'flatness' of the trapezoidal lag-window $\lambda(\frac{s}{M})$ for $|s| < \bar{m}$.

Theorem 2 *Let w be any point in $[-\pi, \pi]$, and 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*

$$Bias(\hat{f}_h(w)) = O(1/M^r). \tag{12}$$

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 we also have

$$Var(\hat{f}_h(w)) = O(M/N). \tag{13}$$

Similarly to the '$2f - f$' rule of equation (4), the estimator $\hat{f}_h(w)$ has very small bias as well, of asymptotic order $O(1/M^r)$. 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. An important class of time series possessing smooth spectral densities is the family of ARMA models; see the discussion after Theorem 3 in Section 3.3.

Note that, under the assumptions of Theorem 2, 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 the $MSE(\hat{f}_h(w))$ is minimized asymptotically by letting $M \sim aN^{1/(2r+1)}$, for some constant $a > 0$, in which case the minimized $MSE(\hat{f}_h(w))$ is 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 (9) 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 3.5.

3.3. 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), 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 $\lambda(\frac{s}{M})$ 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 zeroes 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 4, the spectral kernel $\Lambda_0(w)$ is plotted (for $M = 40$), while in Figure 5, 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 6.

Nevertheless, the good asymptotic performance of $\hat{f}_h(w)$ as demonstrated in Theorem 2 shows that, at least for large samples, the probability of $\hat{f}_h(w)$ being negative is negligible. As far as small samples are concerned, the following immediate modification of $\hat{f}_h(w)$ is proposed to yield a surely nonnegative spectral estimator. Define

$$\hat{f}_h^+(w) = \max(\hat{f}_h(w), 0), \tag{14}$$

i.e., $\hat{f}_h^+(w)$ is the positive part of $\hat{f}_h(w)$. In the following theorem it is shown that taking the positive part results in a better (with respect to MSE) estimator.

Theorem 3 *Let w be any point in $[-\pi, \pi]$. Then, $MSE(\hat{f}_h^+(w)) \le MSE(\hat{f}_h(w))$.*

It now follows that, under the assumptions of Theorem 2, the estimator $\hat{f}_h^+(w)$ is also consistent for $f(w)$, and has the desirable property of being nonnegative. Furthermore, 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 as well, i.e., $MSE(\hat{f}_h^+(w)) = O(N^{-2r/(2r+1)})$. In particular, if f is smooth enough, (i.e., if r as defined in the assumptions of Theorem 1 or 2 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 $\sqrt{N}$.

A comparison of the performance of $\hat{f}_h^+(w)$ to the performance of the usual nonparametric estimators possessing characteristic exponent equal to 2 is in order. Each of the latter results to an estimate with bias of order $O(1/M^2)$, 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

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_n X_{t-n} = Z_t + \psi_1 Z_{t-1} + \cdots + \psi_m Z_{t-m}, \tag{15}$$

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, in which case $\sum_{s=-\infty}^{\infty} |s|^r |R(s)| < \infty$, for *any* positive integer r.

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., $\{X_t\}$ satisfies equation (15) 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.

3.4. 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 the constant value $M = m$.*

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 $\sqrt{N/M} << \sqrt{N}$.

Note however 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 3, $\hat{f}_h^+(w)$ 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 heuristically follows from equation (12) of Theorem 2 with $r = \infty$.

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 $m \leq \bar{m} \leq M$. 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 $\hat{f}(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}k(\frac{s}{M})\hat{R}(s)e^{-jsw}$, with lag-window $k(\frac{s}{M}) = 0$, 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 if 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.

3.5. 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 3.2, 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 $\hat{f}_h^+(w)$ will be very accurate, i.e., the error in the approximation of $f(w)$ by $\hat{f}_h(w)$ or by $\hat{f}_h^+(w)$ will be of order $O(1/\sqrt{N})$ 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.

4. Comments and conclusions

In Sections 2 and 3, an intuitive proposal for bias-corrected nonparametric spectral estimators was presented 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 $2\pi\bar{T}_N(0)$, which is a constant multiple of the Bartlett spectral estimator evaluated at point 0, is a consistent estimator of $Var(\sqrt{N}\bar{X}_N)$; more precisely, $E\left(2\pi\bar{T}_N(0) - Var(\sqrt{N}\bar{X}_N)\right)^2 \to 0$, as $N \to \infty$. As a matter of fact, the estimator $2\pi\bar{T}_N(0)$ 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,b)); 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.

A criticism that might be raised is that the results offered in this paper are asymptotic. For example, it was proved that the variance of $\hat{f}_h(w)$ is $O(M/N)$, similarly to the variance of $\tilde{f}(w)$, but it is expected that the proportionality constants are different. This is in fact true, and is intuitively due to the bias-variance trade-off in smoothing. The finite-sample performance of $\hat{f}_h(w)$ deserves further study and will be investigated by simulation in a follow-up report.

It should be mentioned that this proportionality constant can actually be calculated explicitly; preliminary results of Pedrosa and Schmeiser (1992) indicate that the large-sample correlation coefficient between $\tilde{f}(w)$ and $\bar{f}(w)$ is equal to $(1+\frac{M-\bar{m}}{2M})\sqrt{\bar{m}/M}$. It follows that $Var(\hat{f}_h(w)) \sim \frac{3h+1}{h+1}Var(\tilde{f}(w)) = \frac{3h+1}{h+1}\frac{2M}{3N}f^2(w)(1+\eta(w))$. For the $2f-f$ rule, i.e., the estimator $\hat{f}_1(w)$, it is seen that $Var(\hat{f}_1(w)) \sim 2Var(\tilde{f}(w)) = \frac{4M}{3N}f^2(w)(1+\eta(w))$. 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 $Var(\hat{f}_1(w)) \sim 2Var(\tilde{f}(w))$, how can it be that $\hat{f}_1(w)$ 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 3.5, 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.

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 (cf. Politis and Romano (1992a,b), 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)), and (by the delta-method) so will $\hat{f}_h^+(w)$. As a matter of fact, use of the delta-method also shows the following interesting result.

Lemma 2 *If $f(w)>0$, then $Var(\hat{f}_h^+(w)) \sim Var(\hat{f}_h(w))$, as $N\to\infty$; if $f(w)=0$, then the asymptotic 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.*

Spectral density estimation is now an almost 50 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) 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 apparent that, if the characteristic exponent of a lag-window exceeds r as defined in the assumptions of our Theorem 2, then the bias of the corresponding spectral estimator is of order $O(1/M^r)$. However, the possible nonpositivity of such estimators was considered to be a major drawback, sufficient to limit consideration to lag-windows with characteristic exponent not greater than two. In Section 3.3 it was shown how this nonpositivity is easily side-stepped without sacrificing the good MSE performance of the estimator.

In the same vein, Parzen (1957a) did not consider estimators with infinite characteristic exponent other than 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 6) exhibits quite prominent positive 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 5) does not exhibit such behaviour.

Furthermore, Parzen (1957a) introduced the family of lag-windows given by

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

where q is the characteristic exponent. 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 $0 < h < \infty$ have an infinite characteristic exponent, and share the same properties that are summarized below.

- $\hat{f}_h^+(w)$ can be computed easily and fast, taking into account that the actual computation of Bartlett's estimator via formula (1) is extremely fast.

- The MSE of $\hat{f}_h^+(w)$ is of very small order provided the function $f(w)$ is smooth, having a number of derivatives.

- The rate of convergence of $\hat{f}_h^+(w)$ is $\sqrt{N}$ if the data are MA(m) or m-dependent.

- Last but not least in importance is that working with $\hat{f}_h^+(w)$ 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 THEOREM 1. Omitted in view of the proof of the more general Theorem 2 below.

PROOF OF LEMMA 1. Parts (a)–(c) follow immediately from representation (11) and the properties of the Fejér kernel (cf. Brockwell and Davis (1991)).

To prove part (d) first note that the Fejér kernel has the properties of a probability density with finite support and, in view of part (c), it is immediate that $\int_{-\pi}^{\pi} w^k \Lambda_0(w)dw = O(1/M)$, as $M \to \infty$; however, we will now show that $\int_{-\pi}^{\pi} w^k \Lambda_0(w)dw \sim b_k/M$, for some constant $b_k \neq 0$, i.e., that this kth moment of the Fejér kernel is of *exact* order $O(1/M)$.

Note that

$$\int_{-\pi}^{\pi} w^k \Lambda_0(w)dw = \int_{-\pi}^{\pi} \frac{w^k}{2\pi} \sum_{r=-M}^{M} (1 - \frac{|r|}{M}) e^{jrw} dw$$

$$= \sum_{r=-M}^{M} (1 - \frac{|r|}{M}) \frac{1}{2\pi} \int_{-\pi}^{\pi} w^k \cos(rw)dw.$$

Assuming $r \neq 0$, it is easily calculated by integration by parts that

$$\int_{-\pi}^{\pi} w^k \cos(rw)dw = \frac{2k\pi^{k-1}(-1)^r}{r^2} - \frac{k(k-1)}{r^2} \int_{-\pi}^{\pi} w^{k-2} \cos(rw)dw.$$

What should be noted is that in general, $\int_{-\pi}^{\pi} w^k \cos(rw)dw = C_k \frac{(-1)^r}{r^2} + O(\frac{1}{r^4})$, where C_k is a constant depending on k only; in particular,

$$\int_{-\pi}^{\pi} w^2 \cos(rw)dw = \begin{cases} 4\pi(-1)^r/r^2 & \text{if } r \neq 0 \\ 2\pi^3/3 & \text{if } r = 0. \end{cases}$$

To complete the computation define

$$L_M = \sum_{r=1}^{M} (1 - r/M)(-1)^r/r^2 = \sum_{r=1}^{M} (-1)^r/r^2 - (1/M) \sum_{r=1}^{M} (-1)^r/r.$$

The second term above is $(1/M)\sum_{r=1}^{M}(-1)^r/r \sim \log 2/M$, since $\sum_{r=1}^{\infty}(-1)^r/r = \log 2$; the first term is $\sum_{r=1}^{M}(-1)^r/r^2 = -\pi^2/12 + O(1/M^2)$. To see the last assertion, write $\sum_{r=1}^{M}(-1)^r/r^2 = -\sum_{r=1}^{M} 1/r^2 + 2\sum_{r=1}^{M/2} 1/(2r)^2$, where for convenience and without loss of generality it was assumed that M is even.

Since $\sum_{r=1}^{\infty} 1/r^2 = \pi^2/6$, it follows that

$$-\sum_{r=1}^{M} 1/r^2 = -\pi^2/6 + \sum_{r=M+1}^{\infty} 1/r^2$$

$$= -\pi^2/6 + \int_{r=M+1}^{\infty} \frac{1}{x^2} dx + O(1/M^2) = -\pi^2/6 + 1/M + O(1/M^2),$$

and that

$$2\sum_{r=1}^{M/2} 1/(2r)^2 = 2^{-1}\sum_{r=1}^{M/2} 1/r^2 == 2^{-1}(\pi^2/6 - \sum_{r=(M/2)+1}^{\infty} 1/r^2)$$

$$= \pi^2/12 - (1/2)\int_{r=(M/2)+1}^{\infty} \frac{1}{x^2} dx + O(1/M^2) = \pi^2/12 - 1/M + O(1/M^2).$$

Adding the two we get that $\sum_{r=1}^{M} (-1)^r/r^2 = -\pi^2/12 + O(1/M^2)$.

It now follows that $L_M = -\pi^2/12 - \log 2/M + O(1/M^2)$; therefore, for the specific case $k = 2$ we have

$$\int_{-\pi}^{\pi} w^2 \Lambda_0(w) dw = 4L_M + \pi^2/3 = -4\log 2/M + O(1/M^2),$$

and it is similarly calculated that in general,

$$\int_{-\pi}^{\pi} w^k \Lambda_0(w) dw = b_k/M + O(1/M^2),$$

where b_k is a nonzero constant depending on k only.

As a consequence of the above calculations, it follows that $\int_{-\pi}^{\pi} w^k \Lambda_h(w) dw = (h+1)b_k/M - hb_k/\bar{m} + O(1/M^2)$. It is now apparent that the choice $h = \bar{m}/(M - \bar{m})$ minimizes asymptotically this kth moment $\int_{-\pi}^{\pi} w^k \Lambda_h(w) dw$, making it of order $O(1/M^2)$, and the lemma is proven.

PROOF OF THEOREM 2. 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

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

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

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

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

To complete the proof of equation (12), 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

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

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

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

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

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

But $\lambda(\frac{s}{M}) = 1 - \frac{s-\bar{m}}{M-\bar{m}}$ 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

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

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, and equation (12) is proven.

From equation (9) it now follows that the estimator $\hat{f}_h(w)$ is a linear combination of the two Bartlett spectral estimators $\tilde{f}(w)$ and $\bar{f}(w)$, the first having variance of asymptotic order $O(M/N)$, and the second having variance of asymptotic order $O(\bar{m}/N) = O(M/N)$. By the Cauchy-Schwarz inequality, the covariance between $\tilde{f}(w)$ and $\bar{f}(w)$ is also of order $O(M/N)$; therefore, $Var(\hat{f}(w)) = O(M/N)$ as well, and the theorem is proven.

Proof of Theorem 3. The proof of Theorem 3 is a consequence of the following general lemma. Note that for the lemma, no assumptions whatsoever are required (independence, stationarity, etc.) regarding the probability structure of the sample.

Lemma 3 *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 Lemma 3. 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 lemma.

Theorem 3 now follows from Lemma 3, 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 4. Note that since $N \to \infty$, we can assume without loss of generality that $N > M$. Recall from the proof of Theorem 2 that

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

where A_2 is of order $O(1/N)$ (see the proof of Theorem 2), and now $A_3 = 0$, since it is assumed that $R(s) = 0$, for $|s| > m$. The term A_1 now can be written as

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

where again it was used that $R(s) = 0$, for $|s| > m$. But $\lambda(\frac{s}{M}) = 1$ 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 $Bias(\hat{f}_h(w)) = O(1/N)$. Now $Var(\hat{f}_h(w)) = O(M/N) = O(1/N)$ by Theorem 2 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 3.

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] Daniell,P.J.(1946), Discussion on Symposium on autocorrelation in time series, *J.Roy.Statist.Soc.Suppl.*, 8, 88-90.

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

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

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

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

[10] 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.

[11] 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.

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

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

[14] Pedrosa, A. and Schmeiser, B. (1992), Personal communication on a forthcoming technical report, Department of Industrial Engineering, Purdue University.

[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, to appear in *Ann. Statist.*, vol. 20, No. 4, December 1992.

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

[17] 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.

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

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

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

[21] 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.

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

[23] Welch, P.D. (1967), The use of the Fast Fourier Transform for Estimation of Spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Electr.*, AU-15, 70.

[24] 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.

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

CAPTIONS FOR FIGURES.

FIGURE 1. True spectral density (the most peaked curve), a smoothed estimate (middle line), and an oversmoothed estimate.

FIGURE 2. The lag-window $\hat{\lambda}(\frac{s}{M})$ for different values of the ratio $\bar{m}/M$.
(a) Case $\bar{m} << M/2$.
(b) Case $\bar{m} >> M/2$.
(c) Case $\bar{m} = M/2$.

FIGURE 3. The general trapezoidal lag-window $\lambda(\frac{s}{M})$.

FIGURE 4. The spectral (Fejér) kernel $\Lambda_0(w)$; case $M = 40$.

FIGURE 5. The spectral kernel $\Lambda_1(w)$; case $M = 40$.

FIGURE 6. The spectral (Dirichlet) kernel $\Lambda_\infty(w)$; case $M = 40$.

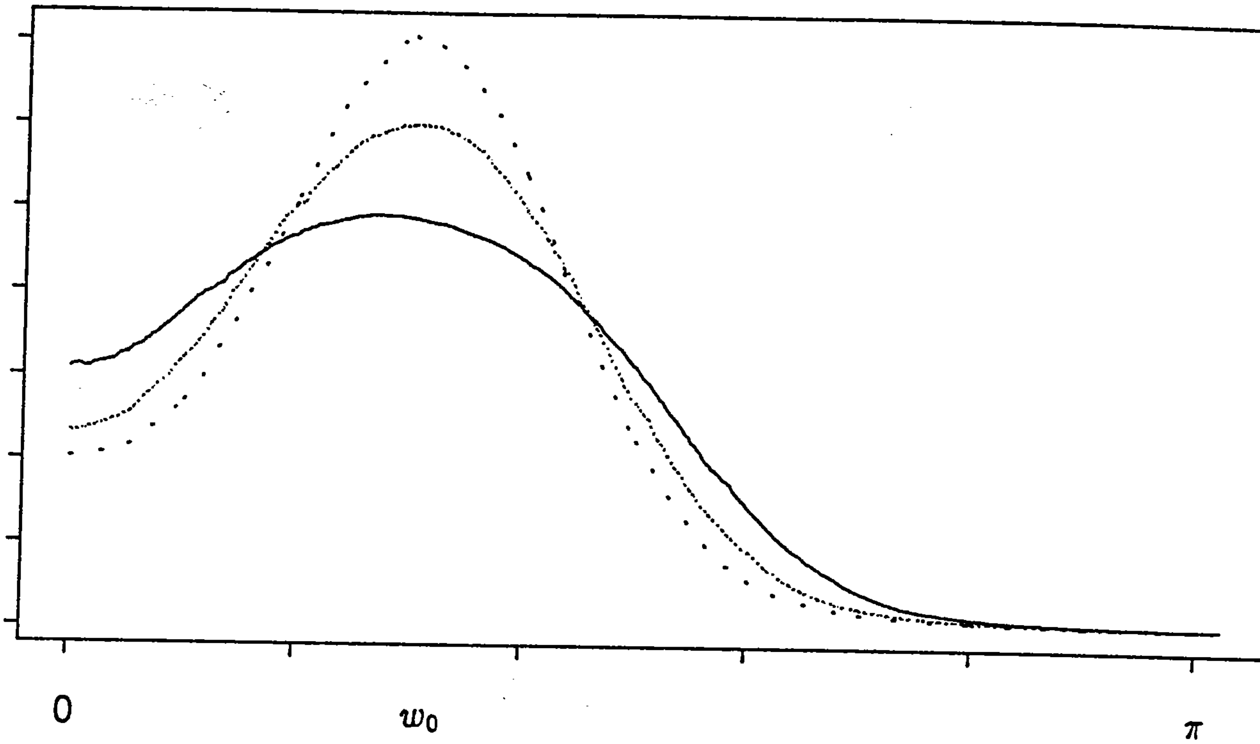

FIGURE 1.

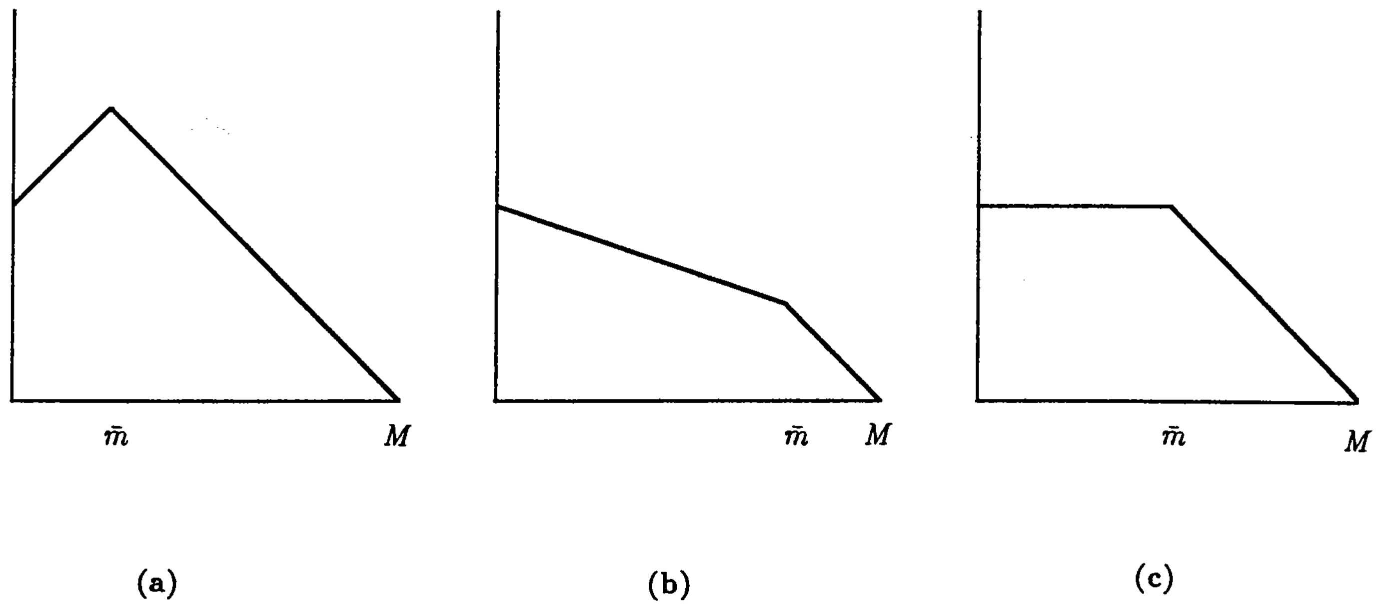

FIGURE 2.

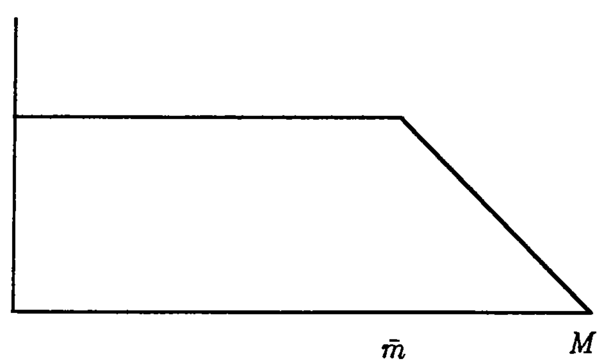

FIGURE 3.

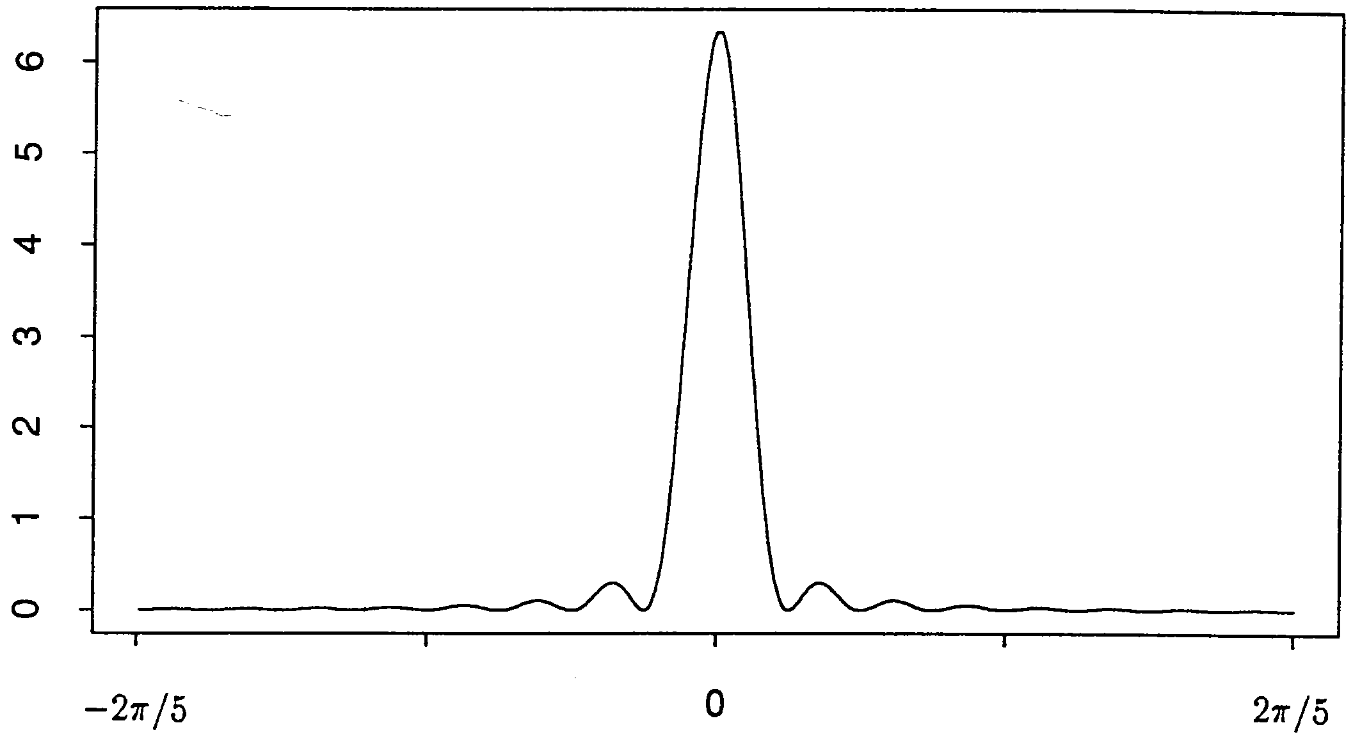

FIGURE 4.

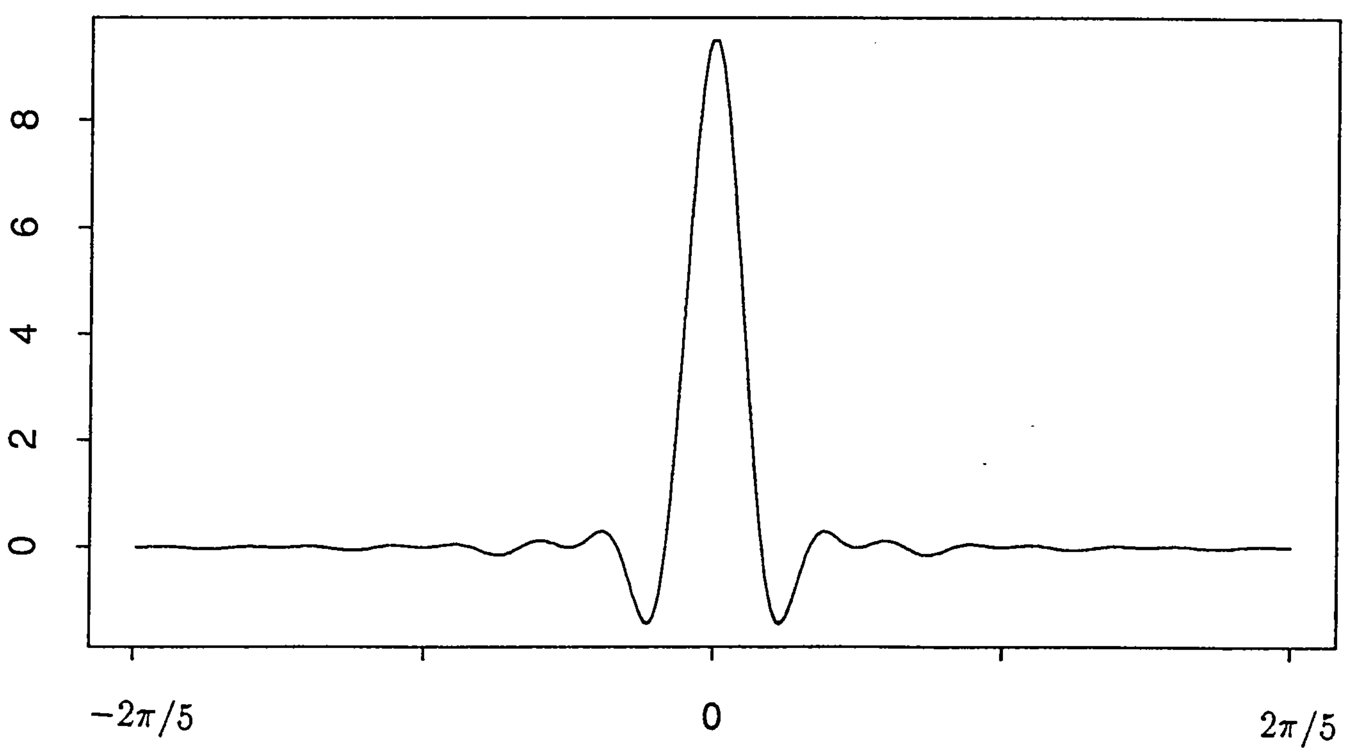

FIGURE 5.

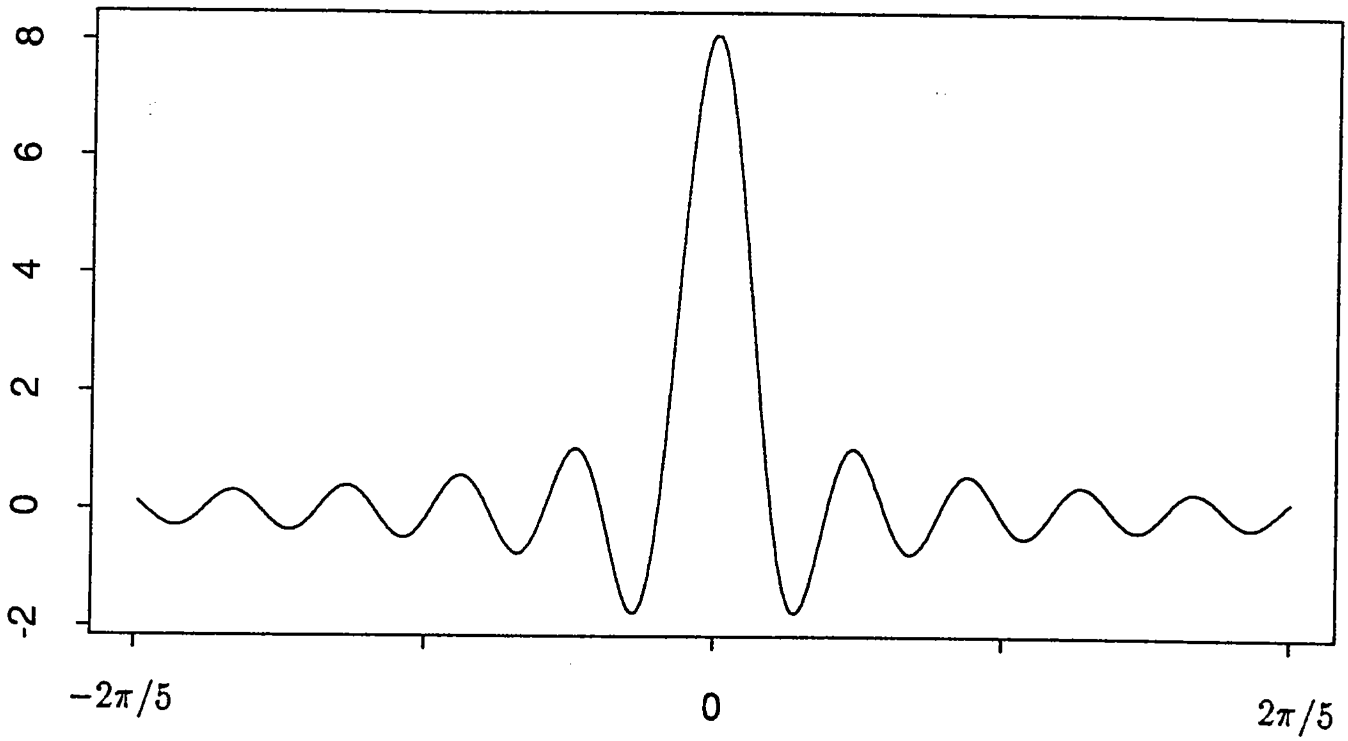

FIGURE 6.