Efficient Nonparametric Spectral Estimation with Varying Bandwidth

Yudianto Pawitan
University of Washington, Seattle

and

Ashis K. Gangopadhyay
University of North Carolina, Chapel Hill

Abstract. We consider an estimation of spectral density at a particular frequency. A linear model is developed based on a weak convergence result expressing a local behaviour of the kernel spectral estimates and a second stage estimate is derived from such model. The estimate is shown to be asymptotically normal and its ARE with respect to the kernel estimate with the optimal, but unknown, bandwidth can be greater than one. A small sample simulation study is conducted which shows good mean square characteristics of the second stage estimate. Practical implementation is discussed through examples.

Keywords. Cumulant mixing, kernel smoothing, optimal bandwidth, weak convergence, spectral estimate.

1. Introduction.

Let X_t be a stationary time series observed at discrete time $t = 0, \pm 1, \pm 2, \ldots$. The second order properties of X_t is equivalently described by the covariance function

$$\gamma(m) = E(X_{t+m}-\mu)(X_t-\mu), \quad m = 0, \pm 1, \pm 2, \ldots$$

and by the spectral density function

$$f(\nu) = \sum_m \gamma(m) \exp(-2\pi\nu m).$$

There is a long history to the problem of estimating $f(\nu)$ from a finite sample $X_0, \ldots, X_{T-1}$ (see, e.g., Priestley, 1981).

The traditional approach has been the nonparametric smoothing of the periodograms or the covariance transform, but, as pointed out by Wahba (1980), the choice of the important smoothing parameter (the window width) is largely left subjective. She then proposed to fit a cross-validated smoothing spline to the log spectrum. Beltrao and Bloomfield (1987) established a cross-validated choice of the bandwidth in kernel spectrum estimation. The approach is global in that it yields one value of bandwidth to use for the whole range of frequency. Cameron (1987) proposed an automatic step function fit to the spectrum based on some optimal partitioning of the frequency band. As he commented the procedure seems to give a good *exploratory* estimate of the spectrum and it does not give a smooth estimate.

The paper is motivated by the approach of Bhattacharya and Mack (1987) and Bhattacharya and Gangopadhyay (1988) in which the problem of choosing a (local) optimal bandwidth is bypassed by carrying out the estimation at several bandwidths and then finding a second stage estimate as a solution of an asymptotic linear model. Section 2 states the conditions and a weak convergence result that motivates the linear model. We develop the estimate in Section 3 and discuss its properties in Section 4. We show that the estimate is asymptotically efficient in mean square relative to the estimate using the optimal bandwidth. This is a good property since the optimal bandwidth is unknown. We then show some simulation results and examples in Section 5.

2. PRELIMINARY RESULT

Let X_t be a zero mean stationary time series observed at $t = 0,1,\dots,T-1$. Define the periodogram at frequency ν as

$$I(\nu) = \frac{1}{T} \left|\sum_t X_t \exp(-2\pi i \nu t)\right|^2 . \tag{1}$$

A smooth periodogram estimate of the spectral density is

$$\hat{f}_T(\nu) = \int_{-.5}^{.5} K_T(\lambda)\, I(\nu-\lambda)d\lambda, \tag{2}$$

where

$$K_T(\lambda) = B_T^{-1}\, K(\lambda\, B_T^{-1}),$$

and $K(\lambda)$ is a symmetric kernel define on $[-.5,.5]$ having bounded first derivative and $\int K(\lambda)d\lambda = 1$. B_T is a smoothing parameter, called the *bandwidth*, and $K(\lambda)$ is called the spectral window. It is well-known that the shape of $K(\lambda)$ is not as important as the size of B_T. For later reference denote

$$\alpha = \int \lambda^2\, K(\lambda)d\lambda. \tag{3}$$

At this point we introduce the following assumptions:

[A1]. The time series X_t is zero mean strictly stationary and cumulant mixing in the sense of Brillinger (1981), i.e.

$$\sum_{a_1 \dots a_{k-1}} |a_j\, \mathrm{cum}(a_1,\dots,a_{k-1})| < \infty$$

for $j=1,2,\dots,k-1$ and $k=2,3,\dots,$ where

$$\mathrm{cum}(a_1,\ldots,a_{k-1}) = \mathrm{cum}(X_t, X_{t+a_1},\ldots,X_{t+a_{k-1}})$$

is the joint cumulant of order k of X_t.

[A2] The spectral density of X_t satisfies

(i) $f(\nu) > 0$

(ii) f'' is continuous at ν.

Simple calculation of the mean square error of (2) (see, e.g., Brillinger, 1981, §5.6) shows that an optimal rate for B_T is of order $T^{-1/5}$, so we consider $B_T = tT^{-1/5}$. Denote $\hat{f}_t(\nu)$ as (2) using $B_T = tT^{-1/5}$ and we will investigate the behaviour of $\hat{f}_t(\nu)$, for fixed ν, as a stochastic process with index t when $T \to \infty$. The symbol $\overset{\mathcal{D}}{\to}$ indicates the weak convergence of distributions.

<u>Theorem 1</u>. Under A1 and A2, for any $0 < a < b < \infty$, as $T \to \infty$ we have

$$T^{2/5}(\hat{f}_t(\nu)-f(\nu)) \overset{\mathcal{D}}{\to} t^{-1}f(\nu)W(t) + \frac{\alpha}{2} f''(\nu)t^2 \tag{4}$$

for $t \in [a,b]$ and $W(t)$ is the standard Brownian motion.

<u>Proof</u>. From (2) and Brillinger (1981)

$$E\ \hat{f}_t(\nu) = B_T^{-1} \int K(\lambda\ B_T^{-1})\ E\ I(\nu-\lambda)d\lambda$$

$$= B_T^{-1} \int K(\lambda\ B_T^{-1})\ (f(\nu-\lambda) + O(T))d\lambda.$$

Under A2, get a second order expansion of f around ν and use (3) to get

$$E\ \hat{f}_t(\nu) = f(\nu) + \frac{\alpha}{2} f''(\nu)\ B_T^2 + o(B_T^2). \tag{5}$$

Now write the LHS of (4) as

$$T^{2/5}(\hat{f}_t-f) = T^{2/5}[(\hat{f}_t-E\ \hat{f}_t) + (E\ \hat{f}_t-f)]$$

$$= T^{2/5}(\hat{f}_t-E\ \hat{f}_t) + \frac{\alpha}{2} f''\ t^2 \tag{6}$$

where we have dropped ν and $o(B_T^2)$ for convenience. For finite number of $t_1,\ldots,t_p$ the joint distribution of

$$\{T^{2/5}(\hat{f}_{t_i}-f) - \frac{\alpha}{2} f'' \, t_i^2\} = \frac{\sqrt{B_T T}}{\sqrt{t_i}} (\hat{f}_t - E\,\hat{f}_t) \overset{\mathcal{D}}{\to} \{\frac{f}{t_i} W(t_i)\} \tag{7}$$

which follows from Brillinger (1981).

The proof is completed by showing the tightness of

$$Y_T(t) \equiv \frac{\sqrt{B_T T}}{\sqrt{t}} \; (\hat{f}_t - E\,\hat{f}_t), \quad t \in [a,b].$$

From Prokhorov (1956, Lemma 2.2) tightness is implied by

$$E|Y_T(s) - Y_T(t)|^A < C|s-t|^B \tag{8}$$

for any $s,t \in [a,b]$ and all T large, where $A > 0$, $B > 1$ and $C > 0$ are constants independent of T. It is straight forward to show from (7) that

$$\sqrt{\frac{st}{|s-t|}} \, (Y_T(t) - Y_T(s)) \overset{\mathcal{D}}{\to} f\, N(0,1). \tag{9}$$

Since the cumulants of $Y_T(t)$ converges, it follows that

$$E|\sqrt{\frac{st}{|s-t|}} \, (Y_T(t) - Y_T(s))|^4 \to E|f\, N(0,1)|^4 = 3f^4. \tag{10}$$

By simple algebra, we can choose any $\epsilon > 0$ and T_0 accordingly such that

$$\begin{aligned} E|Y_T(t) - Y_T(s)|^4 &< \frac{(3f^4+\epsilon)}{(st)^2} \, |s-t|^2 \\ &< a^{-4} \, (3f^4+\epsilon)|s-t|^2 \end{aligned} \tag{11}$$

for all $T > T_0$ and any $s,t \in [a,b]$. Thus (8) is satisfied and the proof is complete.

3. THE PROPOSED ESTIMATOR

Proceeding as in Bhattacharya and Mack (1987) we note that the weak convergence result in Theorem 1 suggests a linear model

$$\begin{aligned} \hat{f}_t &\approx f + \frac{\alpha}{2} f''(t\, T^{-1/5})^2 + f\, e_t \\ &= f + \frac{\alpha}{2} f''\, B_T^2 + f\, e_t \end{aligned} \tag{12}$$

for every $t \in [a,b]$ and where $e_t \overset{\mathcal{D}}{=} t^{-1} T^{-2/5} W(t)$. Clearly $E\, e_t = 0$ and $\mathrm{Cov}(e_{t_1}, e_{t_2}) = T^{1/5} \min(t_1^{-1}, t_2^{-1})$. The best linear estimate of the intercept in (12) is our second stage estimate of f. We will now develop this estimate in a practical form.

It was convenient to prove the result in Section 2 using the continuously smoothed periodogram (2), but in practice $I(\lambda)$ is computed at the discrete Fourier frequencies of the form k/T, $k=0,1,\ldots,[T/2]$, using the fast Fourier transform algorithm. Denote $L = [B_T T] = [tT^{4/5}]$, $0 < a \le t \le b$, $L_0 = [aT^{4/5}]$ and $L_1 = [bT^{4/5}]$, so L takes values $L_0, L_0+1, \ldots, L_1$.

We will consider the rectangular window, hence (2) becomes the simple averaged periodogram

$$\hat{f}_L(\nu) = \frac{1}{L} \sum_{\ell} I\left(\frac{\ell}{T}\right), \tag{13}$$

where the summation is over ℓ satisfying $\left|\frac{\ell}{T} - \nu\right| \le B_T/2$. B_T, then, is approximately L/T. Asymptotic equivalence, in almost sure sense, between discrete and continuous smoothing is stated in Brillinger (1981, Thm. 5.9.1).

The linear model (12) becomes

$$\hat{f}_L = f + \beta(L/T)^2 + f\, e_L, \tag{14}$$

where $L = L_0, L_0+1, \ldots, L_1$, and $\beta = \frac{\alpha}{2} f''$. Now, $Ee_L = 0$ and $\mathrm{cov}(e_{L'}, e_{L''}) = \min(1/L', 1/L'')$.

The second stage estimate of f is the BLUE of f in (14), which is

$$\hat{\hat{f}} = \hat{f}_{L_1} - \hat{\beta}(L_1/T)^2, \tag{15}$$

where $\hat{\beta}$ will be derived below and is shown on (20).

Define

$$\Delta_L = (L(L+1))^{1/2} (e_{L+1} - e_L), \quad L_0 \le L \le L_1 - 1,$$

and

$$\Delta_{L_1} = \sqrt{L_1}\, e_{L_1}.$$

Thus $\{\Delta_L,\ L_0 \le L \le L_1\}$ are mutually uncorrelated with mean 0 and variance 1. From (14) we define the following

$$V_L = (L(L+1))^{1/2} (\hat{f}_{L+1} - \hat{f}_L)$$

$$= \beta\, u_L + f\, \Delta_L, \quad L_0 \le L \le L_1 - 1, \tag{16}$$

$$V_{L_1} = \sqrt{L_1}\, \hat{f}_{L_1}$$

$$= \sqrt{L_1}\, \hat{f}_{L_1} + \beta\, u_{L_1} + f\, \Delta_{L_1} \tag{17}$$

where

$$u_L = (L(L+1))^{1/2} (2L+1)/T^2, \quad L_0 \le L \le L_1 - 1, \tag{18}$$

$$u_{L_1} = L_1^{5/2}/T. \tag{19}$$

The BLUE of β in the linear model defined by (16)-(19) is

$$\hat{\beta} = \left[\sum_{L=L_0}^{L_1-1} u_L V_L\right] \Big/ \left[\sum_{L=L_0}^{L_1-1} u_L^2\right]. \tag{20}$$

4. ASYMPTOTIC PROPERTIES OF $\hat{\hat{f}}$

To study the asymptotic properties of $\hat{\hat{f}}$, note that

$$\sum_{L=L_0}^{L_1-1} u_L^2 = T^{-4} \sum_L L(L+1)(2L+1)^2$$

$$= \int_a^b 4t^4\, dt + O(T^{-4/5})$$

$$\to \frac{4}{5}(b^5 - a^5) \tag{21}$$

and

$$\sum_L u_L V_L = T^{-2} \sum_L L(L+1)(2L+1)(\hat{f}_{L+1}-\hat{f}_L)$$

$$= T^{2/5} [-2a^3 \hat{f}_{L_0} + 2b^3 \hat{f}_{L_1}$$

$$- 6 \int_a^b \{T^{-4/5}[tT^{4/5}]\}^2 \hat{f}_t \, dt] + 0_p(T^{-2/5})$$

$$\overset{\mathcal{D}}{\rightarrow} \frac{4}{5} (b^5-a^5) \beta + f[-2a^2 W(a) + 2b^2 W(b)$$

$$- 6 \int_a^b t W(t)dt] \quad (22)$$

by vitue of the weak convergence result in Theorem 1. Now, substituting (21) and (22) in (15) and (20) we get the following.

Theorem 2: Under the regularity conditions A1 and A2.

(i) a locally asymptotically best linear unbiased estimate of the spectral density f is given by

$$\hat{\hat{f}} = \hat{f}_{L_1} - \hat{\beta}(L_1/T)^2 \quad (23)$$

where

$$\hat{\beta} = \left(\sum_{L=L_0}^{L_1-1} u_L V_L\right) / \left(\sum_{L=L_0}^{L_1-1} u_L^2\right) \quad (24)$$

where u_L and V_L are defined in (16)-(19).

(ii) We also have

$$\{T^{2/5}(\hat{\hat{f}}-f), (\hat{\beta}-\beta)\} \overset{\mathcal{D}}{\rightarrow} f(\theta,\eta), \quad (25)$$

where (θ,η) is bivariate normal with mean (0,0) and covariance

$$\begin{bmatrix} (A+1)/b & -A\, b^{-3} \\ -A\, b^{-3} & A\, b^{-5} \end{bmatrix} \quad (26)$$

where $A = \frac{5}{4} (1-(a/b)^5)^{-1}$.

Remark 1. Note that by Theorem 1, the asymptotic MSE (AMSE) of $\hat{f}_t$ is

$$AMSE(t) = T^{-4/5}[\beta^2 t^4 + f^2/t] \tag{27}$$

and it is minimized at

$$t_0 = [f^2/4\beta^2]^{1/5} \tag{28}$$

So, if we define the asymptotic relative efficiency (ARE) of a given estimator of f with respect to $\hat{f}_{t_0}$ as the ratio of AMSE of $\hat{f}_{t_0}$ to that of the given estimator, then ARE of $\hat{\hat{f}}$ is given by

$$\frac{5}{4}\, t_0^{-1}/\{(A+1)/b\}$$

$$= b\, t_0^{-1}[\frac{4}{5} + \{1-(a/b)\}^{-1}] \tag{29}$$

using theorem 2 (ii). This suggests that we choose b sufficiently large so as to make ARE large. However the choice of b is limited since $L_1 = [bT^{4/5}] < T$ and by the fact that we only use second order approximation.

Remark 2. It should be noted that the second term of the estimator $\hat{\hat{f}}$ in (23) attempts to correct the bias of the first stage estimate at those frequencies where f" is large.

Remark 3. A modification is necessary near the boundary of the frequency range.

5. SIMULATION AND EXAMPLES

In this section we will discuss some properties of $\hat{\hat{f}}$ by way of a Monte Carlo simulation of an ARMA model. It will also address the practical question of choosing L_1. The example shows that as L_1 is large $\hat{\hat{f}}$ achieves comparable MSE properties to those of the simple averaged periodograms with optimal (but, in practice, unknown) smoothing parameter L. In agreement with Remark 1 this suggests that L_1 be chosen large enough such that $\hat{f}_{L_1}$ oversmooths but $\hat{\hat{f}}$ still shows a stable appearance over successive L_1. This point will be elaborated later.

(a) ARMA (2,2) model.

We generate $x_t = -.25\, x_{t-1} - .5\, x_{t-2} + e_t - e_{t-1} - .75\, e_{t-2}$, for $t = 0,1,\ldots,255$ where e_t are iid $N(0,1)$. This model is used in Beamish and Priestley (1981). The simulation is done in GAUSS programming language (1986) which uses the fast acceptance-rejection algorithm proposed by Kinderman and Ramage (1976) to generate the white noise series. The process starts from zero and the first 200 points are dropped so as to get stationarity. The results are based on 100 replications. We use L_0=10 and consider various values for L_1.

Fig. 1

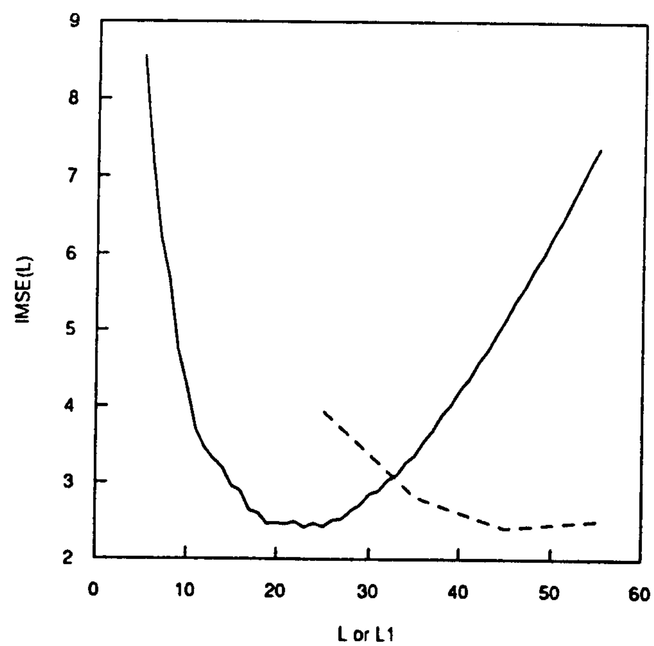


Fig. 1. Integrated MSE (IMSE) of simple averaged periodogram (solid line) as a function of L and IMSE of $\hat{\hat{f}}$ as a function of L_1 (dashed line).

Figure 1 shows the integrated MSE (IMSE) of the simple averaged periodogram, calculated by

$$\mathrm{IMSE}(L) = \frac{1}{33} \sum_{k=1}^{33} \{\frac{1}{100} \sum_{j=1}^{100} (\hat{f}_{L_j} (\tfrac{k-1}{64}) - f(\tfrac{k-1}{64}))^2\} \tag{39}$$

where f is the true spectrum. The corresponding IMSE (L_1) for $\hat{\hat{f}}$ is compared similarly. The summands in (30) are the pointwise MSE. From Figure 1 we see that the optimal L is around 20, which would not be obvious from a particular realization. Notice that as L_1 gets large the IMSE of $\hat{\hat{f}}$ becomes close to optimal. Hence choosing L_1 is relatively simple.

Figure 2 (a), (b) and (c) show the pointwise characteristics of $\hat{\hat{f}}$ in terms of MSE, variance and bias, compared to those of $\hat{f}_L$ with $L=20$ which is known to be optimal from Figure 1. Figure 3 (a), (b), (c) and (d) show $\hat{\hat{f}}$ and $\hat{f}_L$ for a particular realization with $T=256$. We use $L_0=10$ and $L_1=35$, 45, 55 respectively in (b), (c) and (d).

The MSE of $\hat{\hat{f}}$ is very close to that of $\hat{f}_{20}$. With large L_1, $\hat{\hat{f}}$ has smaller MSE around the spectral peak. With moderate L_1, $\hat{\hat{f}}$ has smaller bias but larger variance compared to $\hat{f}_{20}$ or $\hat{\hat{f}}$ with larger L_1. Hence $\hat{\hat{f}}$ can achieve comparable MSE as the optimal simple averaged periodogram, but with a different bias-variance structure.

The example with a particular realization in Figure 3 illustrates the point that L_1 is to be chosen large enough but $\hat{\hat{f}}$ still shows a stable appearance. It is worth noting that the estimate stays similar for a wide range of L_1, but it gets smoother. The optimal $\hat{f}_{20}$ does not look very smooth, which would temp one to try a larger L and, consequently, oversmooth.

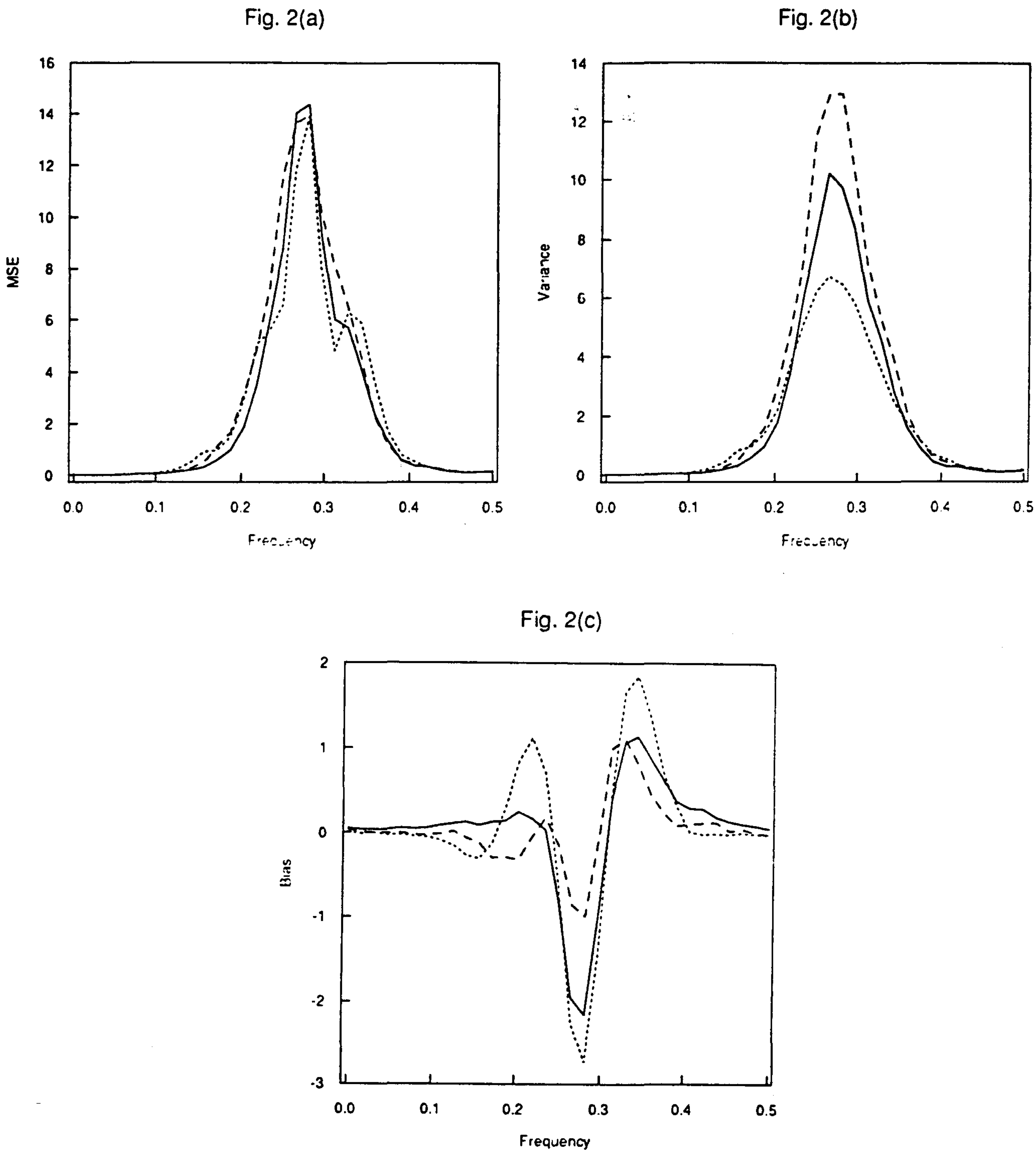


Figure 2. Pointwise characteristics of $\hat{\hat{f}}$ with L_0=10 and L_1=35 (dashed lines), L_1=55 (dotted lines) compared to that of the optimal simple averaged periodogram $\hat{f}_{20}$ (solid lines). (a) MSE, (b) Variance and (c) Bias.

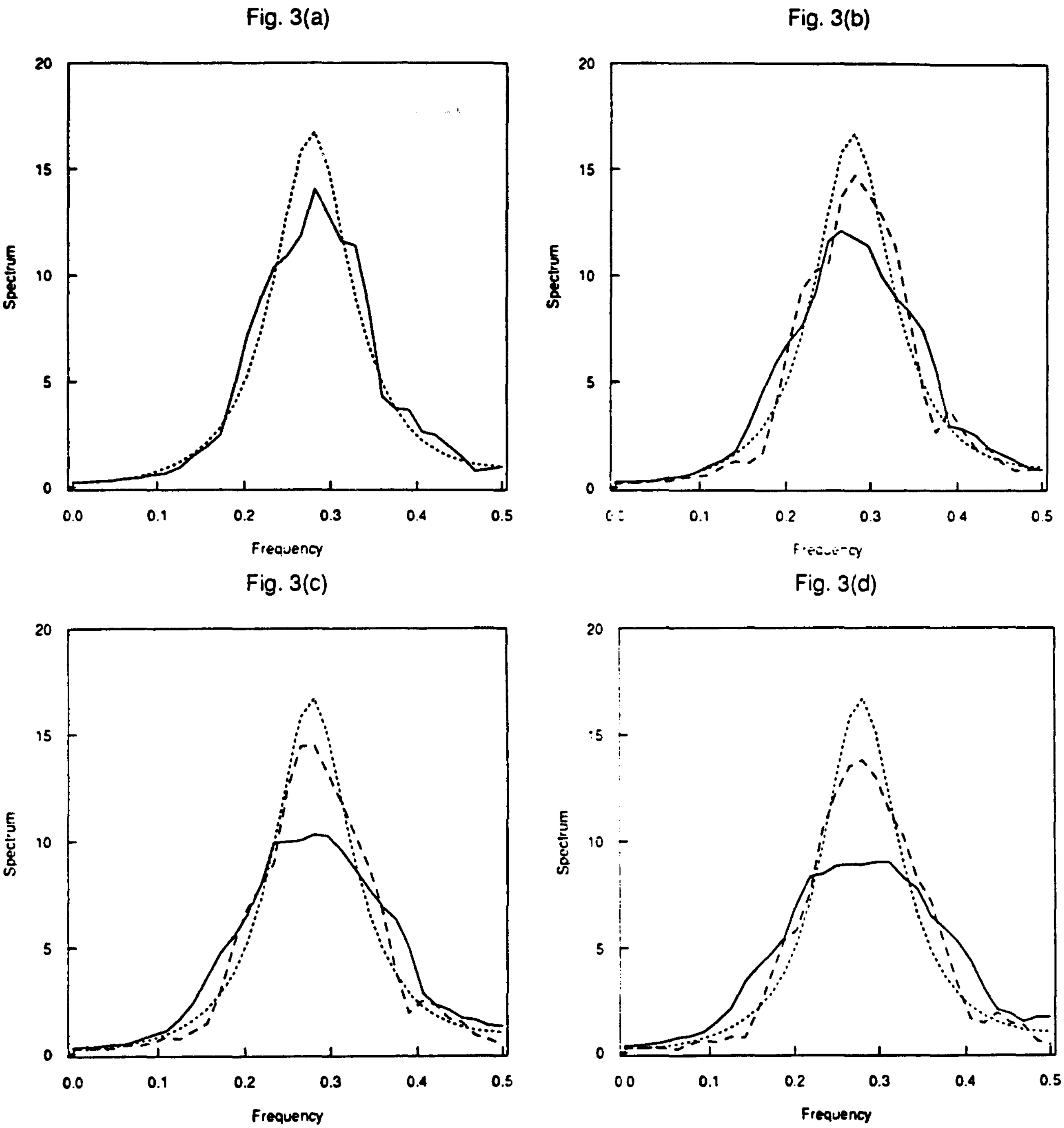


Figure 3. Spectral estimates from a realization of $x_t = -.25\, X_{t-1} - .5\, X_{t-2} + e_t - e_{t-1} - .75\, e_{t-2}$, $t=0,\ldots,255$. The plots show the true spectrum (dotted line) with (a) the optimal $\hat{f}_{20}$ (solid line); (b), (c) and (d) show $\hat{\hat{f}}$ (dashed lines) using $L_0=10$ and $L_1=35$, 45 and 55 respectively, and its corresponding $\hat{f}_{L_1}$ (solid lines).

(b) Other examples.

It is obvious that the choice of L_1 is limited by the range of frequency where the spectrum has a relatively constant second order derivative. Hence a blind choice of large L_1 will not work for a spectrum that has a lot of peaks. A thoughtful procedure like prefiltering is still useful especially to handle sharp peaks. We will show two more examples to see how $\hat{\hat{f}}$ works in more difficult situations than (a).

Figure 4 (a) and (b) show the true spectrum of an AR(4) and $\hat{\hat{f}}$ with $L_0=5$ and $L_1=10$ and 15 respectively. With $L_1=15$, $\hat{f}_{L_1}$ oversmooths the periodogram, so it is probably a good choice for $\hat{\hat{f}}$. This process is considered in Beamish and Priestley (1981). With larger L_1, $\hat{\hat{f}}$ becomes erratic. It should be noted that it is possible for $\hat{\hat{f}}$ to be negative for some value of L_1 and some frequency. This will be an indication that the particular L_1 is not appropriate for the frequency.

Figure 5 (a) and (b) show the true spectrum of an AR(6) and $\hat{\hat{f}}$ with $L_1=15$ and 20 respectively. This example is considered in Lysne and Tjosteim (1987). With $L_1=20$, $\hat{\hat{f}}$ is reasonably good at handling the peaks.

Fig. 4(a) Fig. 4(b)

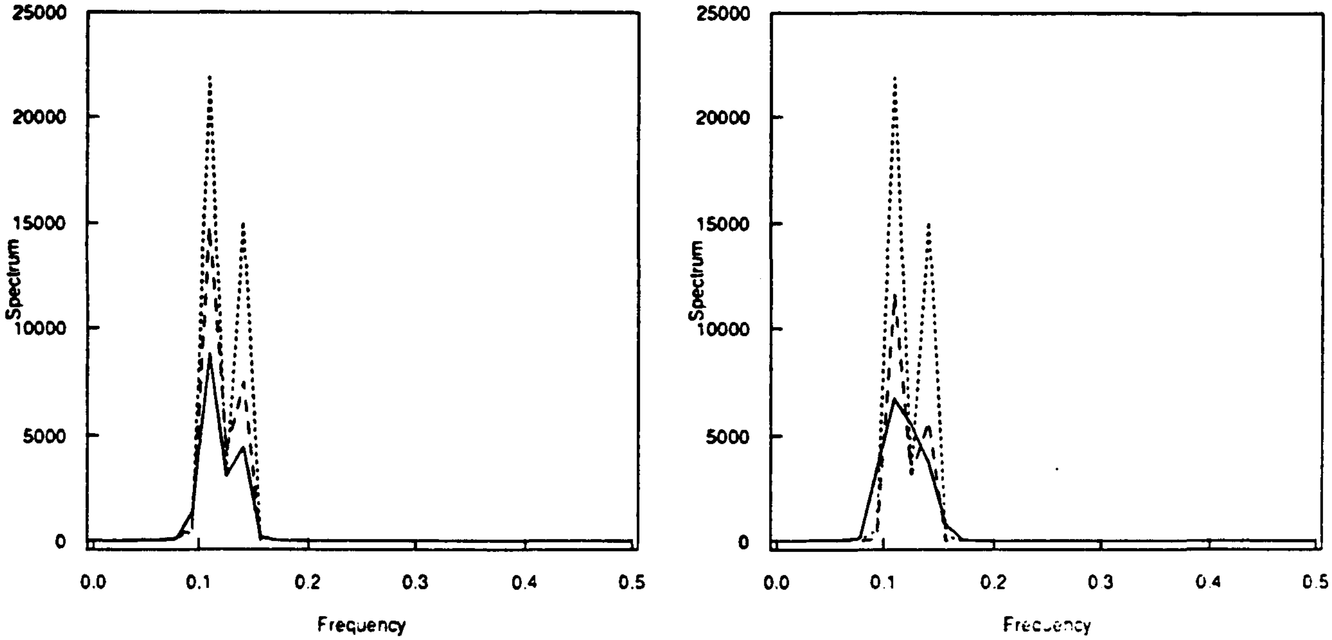
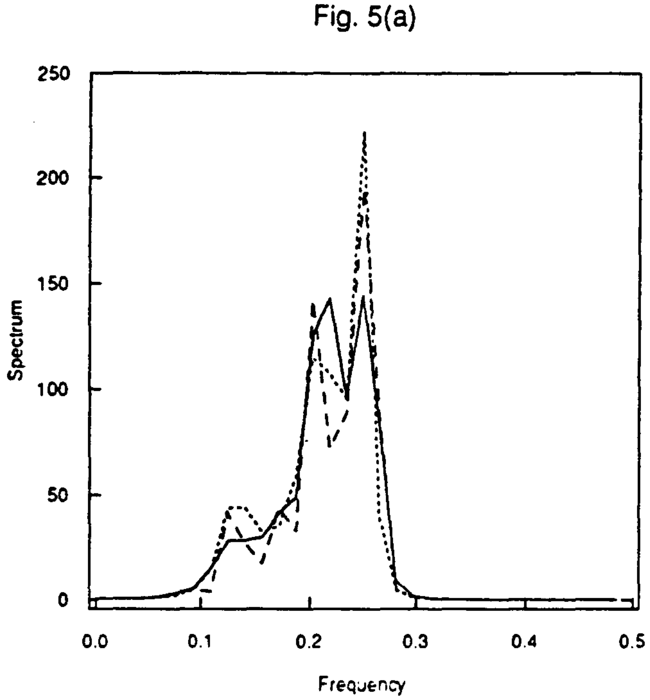


Figure 4. $x_t = 2.7607\, x_{t-1} - 3.8106\, X_{t-1} + 2.6535\, X_{t-3} - .9238\, X_{t-4} + e_t$, $t=0,\ldots,511$. The plot shows the true spectrum (dotted lines), $\hat{f}_{L_1}$ (solid lines) and $\hat{\hat{f}}$ (dashed lines) with $L_0=5$ and (a) $L_1=10$; (b) $L_1=15$.



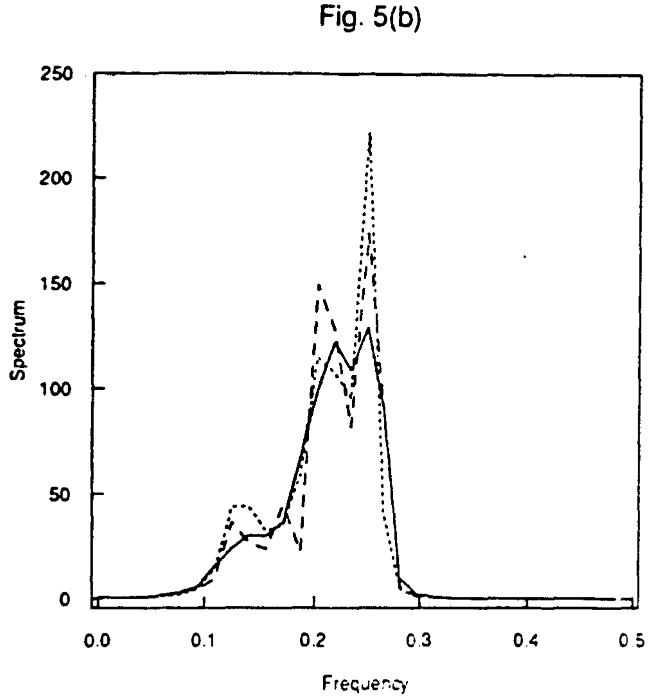


Figure 5. $X_t = 1.68\, X_{t-1} - 3.03\, X_{t-2} + 2.85\, X_{t-3} - 2.59\, X_{t-4} + 1.23\, X_{t-5} - .59\, X_{t-6} + e_t$, $t=0,\ldots,511$. The plot shows the true spectrum (dotted lines), $\hat{f}_{L_1}$ (solid lines) and $\hat{\hat{f}}$ (dashed lines) with $L_0=5$ and (a) $L_1=15$; (b) $L_1=20$.

6. CONCLUSION

We have considered the problem of estimating spectral density functions using a two-staged procedure. It is motivated by the local behavior of kernel estimates with varying bandwidths. The estimate is shown to achieve good mean square characteristics asymptotically and in the simulation, even compared to the kernel estimate with optimal bandwidth, thus reducing the need to find such optimal bandwidth.

REFERENCES

BEAMISH, N. and PRIESTLEY, M.B. (1981). A study of autoregressive and window spectral estimation. *Appl. Statist.* 30, 41-58.

BELTRAO, K.I. and BLOOMFIELD, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Series Anal.* 8, 21-38.

BHATTACHARYA, P.K. and MACK, Y.P. (1987). Weak convergence of k-NN density and regression estimators with varying k and applications. *Ann. Statist.* 15, 976-994.

BHATTACHARYA, P.K. and GANGOPADHYAY, A.K. (1987). *Technical Report Series of the Intercollege Division of Statistics, University of California at Davis, Technical Report #104.*

BRILLINGER, D.R. (1981). *Time Series Data Analysis and Theory.* New York: Holt, Rinehart and Winston.

CAMERON, M.A. (1987) An automatic non-parametric spectrum estimation. *J. Time Series Anal.* 8, 379-387.

GAUSS. Programming Language Manual (1986). Aptech Inc.

LYSNE, D. and TJOSTEIM, D. (1987). Loss of spectral peaks in autoregressive spectral estimation. *Biometrika* 74, 200-6.

KINDERMAN, A.J. and RAMAGE, J.G. (1976). Computer generation of normal random numbers. *J. Amer. Statist. Assoc.* 71, 893-6.

PRIESTLEY, M.B. (1981). *Spectral Analysis and Time Series.* New York: Academic Press.

PROKHOROV, Y.V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Prob. Applic.* 1, 157-214.

WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* 75, 122-132.

Department of Biostatistics
University of Washington
Seattle, WA 98105

Department of Statistics
University of NC at Chapel Hill
Chapel Hill, NC 27599-3260