Ann. Inst. Statist. Math.
Vol. 56, No. 4, 611–630 (2004)

SPECTRAL DENSITY ESTIMATION WITH AMPLITUDE MODULATION AND OUTLIER DETECTION*

JIANCHENG JIANG[1] AND Y. V. HUI[2]

[1] *LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China,* e-mail: jiang@math.pku.edu.cn
[2] *Department of Management Sciences, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China,* e-mail: msyervan@cityu.edu.hk

(Received October 31, 2002; revised October 14, 2003)

Abstract. This paper studies spectral density estimation based on amplitude modulation including missing data as a specific case. A generalized periodogram is introduced and smoothed to give a consistent estimator of the spectral density by running local linear regression smoother. We explore the asymptotic properties of the proposed estimator and its application to time series data with periodic missing. A simple data-driven local bandwidth selection rule is proposed and an algorithm for computing the spectral density estimate is presented. The effectiveness of the proposed method is demonstrated using simulations. The application to outlier detection based on leave-one-out diagnostic is also considered. An illustrative example shows that the proposed diagnostic procedure succeeds in revealing outliers in time series without masking and smearing effects.

Key words and phrases: Amplitude modulation, local linear regression, missing observations, outlier detection, spectral density.

1. Introduction

Let $\{Y_t, t = 0, \pm 1, \pm 2, \dots\}$ be a stationary time series with mean zero and autocovariance function $R_Y(v) = E[Y_t Y_{t+v}]$, $v = 0, \pm 1, \pm 2, \dots$. Then the periodogram for the observed time series $y_1, \dots, y_n$ is given by

$$(1.1) \qquad I_Y^{(n)}(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} y_t \exp(-it\omega) \right|^2, \qquad \omega \in [0, \pi].$$

It is well-known that the periodogram is an asymptotically unbiased estimator of the spectral density function

$$(1.2) \qquad f_Y(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} R_Y(t) \exp(-it\omega), \qquad \omega \in [0, \pi],$$

*Supported by Chinese NSF Grants 10001004 and 39930160, and Fellowship of City University of Hong Kong.

even though it is inconsistent. See for example Brillinger (1981), Priestley (1981) and Brockwell and Davis (1991). If $\{Y_t\}$ follows a stationary linear process, then the periodogam $I_Y^{(n)}(\omega_k)$ at Fourier frequencies $\omega_k = 2\pi k/n$ (for $k = 0, 1, \ldots, N$, where $N = [(n-1)/2]$) are asymptotically independent. The asymptotic unbiasedness and independence of the periodogram allow one to construct a consistent estimator of $f_Y(\omega)$ by locally averaging the periodogram. Most traditional methods are based on this approach. See for example Brillinger (1981), Fan and Kreutzberger (1998), and others.

Other alternative estimators, such as the smoothed log-periodogram and Whittle likelihood-based estimator, have received much attention. Examples include: the automatic smoothing of the log-periodogram in Wahba (1980); the penalized maximum likelihood spline estimator in Pawitan and O'Sullivan (1994); the logspline estimator in Kooperberg *et al.* (1995*a*, 1995*b*); and the local Whittle's likelihood estimator in Fan and Kreutzberger (1998). As demonstrated by Fan and Kreutzberger (1998), the Whittle likelihood-based estimators are desirable and outperform other estimators at regions where the log-spectral density is convex. Fan and Kreutzberger (1998) also recommend using Whittle's likelihood-based method.

However, our experience shows that there is little advantage of using Whittle's likelihood over the smoothed periodogram. One reason is that its asymptotic bias depends on the spectral density and the second derivative of log-spectral density, which complicates the choice of optimal bandwidth selector. Another reason is that the absolute value of the asymptotic bias of the Whittle-based estimator is larger than that of the smoothed periodogam in the region where $f_Y(\omega)$ is convex, while the asymptotic variances are the same (see Remark 2 in Fan and Kreutzberger (1998) for bias comparison). Among these competitive approaches in spectral density estimation, both the smoothed periodogram and the Whittle likelihood-based estimator are asymptotically efficient. They also outperform the smoothed log-periodogram (see Fan and Kreutzberger (1998)).

When there are missing observations, direct use of the aforementioned methods is infeasible because the definition of periodogram in (1.1) is unclear. In addition, these approaches are nonrobust against outliers since they are based on least-squares or local maximum likelihood principles. Several authors have proposed regarding the missing observations or outliers as zeros, and then estimating the spectral density from the "zeroed" data series. See for example Jones (1962), Parzen (1961, 1963), Priestley (1981), and Hui and Lee (1992) among others. A remarkable work in this field is that of Parzen (1963) in which amplitude modulation mechanics is proposed. We will employ Parzen's mechanics to develop a generalized periodogram and extend the smoothed periodogram of Fan and Kreutzberger (1998) in this study. The asymptotic normality of the proposed estimator of spectral density will be proven. The technical derivations of such asymptotic property is very involved and determined efforts have been made. For robustness, one may smooth the logarithm of the generalized periodogram and then estimate the spectrum using an exponential transformation. The proposed approach can be adapted naturally to the smoothed log-periodogram.

As pointed out in Fan and Gijbels (1996), the spectral densities are usually very rough and the periodogram is highly heteroscedastic. Global bandwidth smoothing is usually unsatisfactory in revealing the complicated structure of the frequently changing spectral density. We will develop a simple data-driven local bandwidth selection method for our spectral density estimator, which facilitates the application in outlier detection.

Note that the presence of outliers can severely distort conventional spectral estimators even though such outliers are not large relative to the scale of the observations (see Kleiner *et al.* (1979)). Outlier detection is vital to nonrobust spectral estimation and the existence of masking and smearing effects in time series further complicates the problem. We study an outlier detection procedure by evaluating the smoothed periodogram based on the widely used leave-one-out diagnostics. Therefore the spectral density can be estimated using the proposed method after outliers are removed from the data.

This paper is organized as follows. Section 2 introduces the generalized periodogram and the spectral density estimator. Section 3 considers asymptotic properties of the generalized periodogram and the spectral density estimator. A data-driven bandwidth selection rule is proposed and an algorithm for calculating the spectral density estimate is presented in Section 4. A simulation study on the performance of the proposed estimator for incomplete series is discussed in Section 5. Section 6 studies the application to outlier detection and Section 7 gives our conclusion. Technical proofs are presented in the Appendix.

2. Estimation

2.1 *Parzen's amplitude modulation mechanics*

Consider the amplitude modulation mechanics for spectral density estimation with missing values in Parzen (1963). The time series we are interested in is $\{Y_t\}$, but the observed series is $\{X_t\}$. The amplitude of time series $\{Y_t\}$ is modulated by $\{g_t\}$ through the relationship

$$X_t = g_t Y_t, \tag{2.1}$$

where $\{g_t\}$ is a non-random bounded series possessing a generalized harmonic analysis in the sense that for $v = 0, 1, \ldots,$

$$R_g(v) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n-v} g_t g_{t+v} \tag{2.2}$$

exists. For a systematically unobservable series, such as a stock index which is unobservable on certain days in a year, it may be more appropriate to model this phenomenon as an amplitude modulated series with g_t defined by

$$g_t = \begin{cases} 0 & \text{if } Y_t \text{ is missing at time } t, \\ 1 & \text{if } Y_t \text{ is observed at time } t. \end{cases} \tag{2.3}$$

For a general missing pattern, (2.1) can also be used by letting X_t represent the observed value of Y_t with 0 inserted in the series whenever the value of Y_t is missing.

2.2 *Generalized periodogram*

Suppose $\{X_t\}$ is asymptotically stationary, i.e. $R_X(v) = \lim_{n\to\infty} 1/n \sum_{t=1}^{n-v} E[X_t X_{t+v}]$ exists. Then (2.1) leads to $R_X(v) = R_g(v) R_Y(v)$. According to Parzen (1963), if $\{Y_t\}$ is an ergodic normal process, then $\{X_t\}$ is ergodic. The sample autocovariance function of $\{X_t\}$, $\hat{R}_X(v) = 1/n \sum_{t=1}^{n-v} X_t X_{t+v}$, is a consistent estimator of $R_X(v)$ in the quadratic mean. Therefore, if $R_g(v) \neq 0$, then $R_Y(v)$ can be consistently estimated by

$$\hat{R}_Y(v) = \frac{\hat{R}_X(v)}{R_g(v)}, \quad \text{for} \quad v = 0, 1, \ldots. \tag{2.4}$$

For $v = -1, -2, \ldots$, let $\hat{R}_Y(v) = \hat{R}_Y(-v)$. This motivates us to define the following generalized periodogram.

DEFINITION 2.1. For the time series $\{Y_t\}$ with modulated amplitude, its generalized periodogram is defined as

$$GI_Y^{(n)}(\omega_k) = \frac{1}{2\pi} \sum_{|v|<n} \hat{R}_Y(v) \exp(-iv\omega_k), \tag{2.5}$$

where $\omega_k = \frac{2\pi k}{n}$ $(k = 0, 1, \ldots, N)$ are Fourier frequencies.

The generalized periodogram has properties similar to the common periodograms. It can be used to construct a variety of tests for hidden periodicities following the ideas of common tests based on the periodogram in (1.1). The generalized periodogram has the following asymptotic representation for the bias and covariances.

PROPOSITION 2.1. *If $\{Y_t\}$ is ergodic and normal with mean zero, then the following results hold for $k = 1, \ldots, N$*

(1) *Asymptotic unbiasedness*: $E[GI_Y^{(n)}(\omega_k)] = f_Y(\omega_k) + O(\frac{1}{\sqrt{n}})$.

(2) *Variance approximation*: $\mathrm{Var}[GI_Y^{(n)}(\omega_k)] = f_Y^2(\omega_k) + \tau_Y^2(\omega_k) + o(1)$.

(3) *Asymptotic independence*: *for* $\omega_j \neq \omega_k$,

$$\mathrm{Cov}(GI_Y^{(n)}(\omega_j), GI_Y^{(n)}(\omega_k)) = O\left(\frac{1}{n}\right).$$

A detailed proof of Proposition 2.1 is given in the Appendix and $\tau_Y^2(\omega)$ is defined in Theorem 3.1.

2.3 *Local linear regression smoother*

Note that the generalized periodogram is an asymptotically unbiased estimator of $f_Y(\omega)$ at Fourier frequencies. However, it is inconsistent in estimating $f_Y(\omega)$. A consistent estimator of $f_Y(\omega)$, for $\omega \in [0, \pi]$, can be obtained via directly smoothing the data $\{(\omega_k, GI_Y^{(n)}(\omega_k)), k = 1, \ldots, N\}$ using a locally weighted average. Following Fan and Kreutzberger (1998), we run the local linear regression smoother to the data, and obtain the estimator

$$\hat{f}_Y(\omega) = \sum_{k=1}^{N} K_N\left(\frac{\omega - \omega_k}{h}\right) GI_Y^{(n)}(\omega_k), \tag{2.6}$$

where

$$K_N(t) = \frac{1}{Nh} \frac{s_{N,2} - ht \cdot s_{N,1}}{s_{N,0} \cdot s_{N,2} - s_{N,1}^2} K(t),$$

where $K(t)$ is a kernel function and $s_{N,\ell} = \frac{1}{Nh} \sum_{k=1}^{N} K(\frac{\omega - \omega_k}{h})(\omega - \omega_k)^\ell$.

Other estimation approaches such as Whittle's likelihood in Fan and Kreutzberger (1998), may be indiscriminately applied, but we will not pursue these options further in this paper. Furthermore, the distribution of $GI_Y^{(n)}(\omega_k)$ is unknown under a general amplitude modulation or missing mechanism. It is unclear whether Whittle's likelihood principle applies in these situations.

3. Asymptotic properties

In this section, we only consider the important case where g_t is a function with period θ and Y_t is a moving average process given by

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad Z_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2). \tag{3.1}$$

Assume that $\sum_j |\psi_j| j^2 < \infty$. The spectral density of $\{Y_t\}$ is

$$f_Y(\omega) = \frac{\sigma^2}{2\pi} |\Psi(e^{-i\omega})|^2,$$

where $\Psi(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\omega)$. It follows that the spectral density has a bounded second derivative. From Parzen ((1963), pp. 388–389), g_t admits a harmonic representation

$$g_t = \sum_{k=-\Theta}^{\Theta} e_k \exp(it\lambda_k) G_k, \tag{3.2}$$

where $\lambda_k = \frac{2\pi k}{\theta}$, $\Theta = [\frac{\theta}{2}]$, $G_k = \frac{1}{\theta}\sum_{s=1}^{\theta} \exp(-is\lambda_k) g_s$ for $k = 0, \pm 1, \ldots, \pm\Theta$, and $e_k = 1$ for all k except that $e_{\pm\Theta} = \frac{1}{2}$ if θ is even. Furthermore,

$$R_g(v) = \sum_{k=-\Theta}^{\Theta} e_k e_{-k} G_k G_{-k} \exp(-iv\lambda_k). \tag{3.3}$$

When g_t is not a periodic function and n is moderately large, one can use the approximation

$$R_g(v) \approx \frac{1}{n} \sum_{t=1}^{n-v} g_t g_{t+v}. \tag{3.4}$$

The following assumptions and notations are needed to derive the asymptotic properties of the proposed estimator:

(i) $\{Y_t\}$ is a stationary and normal process with $\sum_j |\psi_j| j^2 < \infty$.

(ii) The spectral density function $f_Y(\omega)$ is positive on $[0, \pi]$.

(iii) The kernel function $K(x)$ is a symmetric probability density function and has a compact support, $[-1, 1]$ say. Let $\mu_2(K) = \int_{-1}^{1} u^2 K(u) du$ and $v_0(K) = \int_{-1}^{1} K^2(u) du$.

(iv) $h \to 0$ and $Nh \to \infty$ as $n \to \infty$.

(v) The bounded amplitude modulated series $\{g_t\}$ with period θ satisfies $\rho = \min_v R_g(v) > 0$ and $\frac{1}{n}\sum_{t=1}^{n-v} g_t g_{t+v} = R_g(v) + O(\frac{1}{\sqrt{n}})$, uniformly in v.

Conditions (i)–(iv) are derived from Fan and Kreutzberger (1998). For function g_t with period θ, $R_g(v) = \frac{1}{\theta}\sum_{t=1}^{\theta} g_t g_{t+v}$ (see (2.4) in Parzen (1963)) and condition (v) follows. An example where these conditions hold is Parzen's periodic missing pattern (see Parzen (1963), pp. 385–386).

THEOREM 3.1. *Suppose assumptions* (i)–(v) *hold. For each* $\omega \in (0, \pi)$, *we have*

$$\sqrt{Nh}\left[\hat{f}_Y(\omega) - f_Y(\omega) - \frac{1}{2} h^2 f_Y''(\omega) \mu_2(K) + o(h^2)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)]).$$

For a boundary point $\omega_n^* = ch$, we have

$$\sqrt{Nh}\left[\hat{f}_Y(\omega_n^*) - f_Y(\omega_n^*) - \frac{1}{2}h^2 f_Y''(0+)\mu_2(K,c) + o(h^2)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, v_0(K,c)\pi[f_Y^2(0+) + \tau_Y^2(0+)]),$$

where

$$\mu_2(K,c) = \frac{s_{2,c}^2 - s_{1,c}s_{3,c}}{s_{0,c}s_{2,c} - s_{1,c}^2}, \qquad v_0(K,c) = \frac{\int_{-1}^{c}(s_{2,c} - s_{1,c}t)^2K^2(t)dt}{(s_{0,c}s_{2,c} - s_{1,c}^2)^2},$$

and

$$\tau_Y^2(\omega) = \sum_{k\neq l} e_k e_l |H_{k,l}|^2 f_Y(\omega+\lambda_k) f_Y(\omega+\lambda_l),$$

with

$$H_{k,l} = \sum_{i=-\Theta}^{\Theta} e_i W_i G_{i-k} G_{-i-l}, \qquad W_j = \frac{1}{\theta}\sum_{s=1}^{\theta} \exp(-is\lambda_j)/R_g(s),$$

and

$$s_{j,c} = \int_{-1}^{c} t^j K(t)dt.$$

Theorem 3.1 gives the asymptotic distribution of the spectral density estimator. A detailed proof is given in the Appendix.

Remark 3.1. It is worth noting that the estimated spectral density inherits the boundary adaptation as discussed in Fan and Kreutzberger (1998). The estimator achieves the same convergence rate at the boundary as at the interior points. When $g_t \equiv 1$, i.e. Y_t is always observed, we have $\tau_Y^2(\omega) = 0$. Our result coincides with the result in Remark 2 of Fan and Kreutzberger (1998). (Note that the convergence rate $\sqrt{nh}$ in Fan and Kreutzberger (1998) is a typo.) Therefore, our result can be regarded as an extension of Fan and Kreutzberger (1998).

Remark 3.2. From Theorem 3.1, the asymptotic mean squared errors (*MSE*) for estimating $f_Y(\omega)$ at $\omega \in (0,\pi)$ can be defined as

$$MSE(h,\omega) = \frac{1}{4}h^4[f_Y''(\omega)\mu_2(K)]^2 + \frac{1}{Nh}v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)]. \tag{3.5}$$

Therefore, the optimal local bandwidth for estimating $f_Y(\omega)$ in the sense of minimizing $MSE(h,\omega)$, is

$$h_{\text{opt}}(\omega) = N^{-1/5}\left(\frac{v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)]}{[f_Y''(\omega)\mu_2(K)]^2}\right)^{1/5}. \tag{3.6}$$

4. Data-driven local bandwidth selection

Identifying peaks of spectral density is of special interest since information on energy and period are carried by the peaks. The global optimal bandwidth works pretty well in the estimation of a generally smooth spectral density. However, the local bandwidth selector should be employed to capture the complicated structure (especially the peaks) of a changing spectral density.

For the nonparametric regression estimator (2.6), common global and local bandwidth choices developed in nonparametric regression literature can be applied. Examples include the "*pre-asymptotic substitution method*" of Fan and Gijbels (1995), the "*plug-in bandwidth selection*" in Ruppert *et al.* (1995), and the "*empirical bias bandwidth selector*" of Ruppert (1997). However, these bandwidth selection methods usually involve a large computational burden for outlier detection. We present a simple rule for bandwidth selection which is easy to implement in practice.

From (A.4)–(A.7) in the Appendix, we know that the exact bias of $\hat{f}_Y(\omega)$ is

$$Bias(h,\omega) \equiv \sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) f_Y(\omega_k) - f_Y(\omega),$$

up to a negligible term of order $O_p(1/\sqrt{n})$. We can evaluate $Bias(h,\omega)$ and the asymptotic variance in Theorem 3.1 for a given bandwidth h at ω if we have a pilot estimate $\hat{f}_Y(\omega)$ for $f_Y(\omega)$ in hand. We propose choosing the bandwidth minimizing the estimated mean squared errors

$$\widehat{MSE}(h,\omega) = [\widehat{Bias}(h,\omega)]^2 + \frac{1}{Nh} v_0(K)\pi[\hat{f}_Y^2(\omega) + \hat{\tau}_Y^2(\omega)], \tag{4.1}$$

where $\widehat{Bias}(h,\omega)$ and $\hat{\tau}_Y^2(\omega)$ are defined as $Bias(h,\omega)$ and $\tau_Y^2(\omega)$ respectively with $f_Y(\omega)$ replaced by $\hat{f}_Y(\omega)$. Define

$$\hat{h}_{\mathrm{opt}}(\omega) = \arg\min_h \widehat{MSE}(h,\omega).$$

Then we minimize the one-dimensional function $\widehat{MSE}(h,\omega)$ with respect to h at each fixed frequency ω. $\hat{h}_{\mathrm{opt}}(\omega)$ can be evaluated at a grid of frequency points. The above bandwidth selection requires only a pilot estimate of the spectral density, which avoids estimation of higher order derivatives of the regression function in implementation of the "*pre-asymptotic substitution*" and "*plug-in*" methods. The proposed method is also simpler than the "*empirical bias bandwidth selector*" method proposed by Ruppert (1997), where the empirical bias is estimated by calculating the estimate of regression function $\hat{m}(x,h)$ on a grid of bandwidth values and then modeling the behavior of $\hat{m}(x,h)$ as h varies. Note that the asymptotic variance in (4.1) does not involve the variance of the white noise in (3.1). It follows that evaluation of the asymptotic variance is much easier compared with the nonparametric regression case where the unknown error variance needs to be estimated from the data.

Common Newton's iterations can be applied to find the $\hat{h}_{\mathrm{opt}}(\omega)$ efficiently. We here synthesize the ideas in minimizing the asymptotic mean squared errors of the hazard rate estimator in Müller and Wang (1994) and maximizing the local pseudo-likelihhod in Fan *et al.* (2003), and propose an algorithm to evaluate $\hat{f}_Y(\omega)$.

Algorithm for estimating $f_Y(\omega)$

Step 1. *Initial estimate of* $f_Y(\omega)$: Choose a kernel, such as the Epanechnikov kernel, and an initial global bandwidth h_0. The choice of the initial bandwidth depends on the specific case. A possible value for h_0 is $\frac{\pi}{8}n^{-1/5}$, as recommended by Müller and Wang (1994), or the global optimal bandwidth estimate given in Fan and Kreutzberger (1998). The initial estimate $\hat{f}_Y(\omega)$ of $f_Y(\omega)$ is obtained by employing $h \equiv h_0$ and (2.6). The pilot estimate of $f_Y(\omega)$ may also be obtained via a parametric approach if one has a plausible parametric model in mind.

Step 2. *Minimization of* $\widehat{MSE}(h,\omega)$: Choose an equispaced grid of m_1 points $\tilde{\omega}_i$, $i = 1,\ldots,m_1$ between 0 and π. For each grid point $\tilde{\omega}_i$, compute $\widehat{MSE}(h,\tilde{\omega}_i)$ in (4.1) and obtain its minimizers $\tilde{h}(\tilde{\omega}_i)$ on the interval, $[h_0/4, 10h_0]$ say. The minimizers can be easily obtained using the estimate $\tilde{h}(\tilde{\omega}_k)$ as the initial value of minimization at the next grid point $\tilde{\omega}_{k+1}$. The minimizer can be found within a few iterations.

Step 3. *Bandwidth smoothing*: Choose another equispaced grid of m_2 points ω_r, $r = 1,\ldots,m_2$, over the interval $[0,\pi]$ on which the final spectral density estimate is desired. Running the following local linear smoother by employing the global bandwidth $\tilde{h}_0 = h_0$ or $2h_0$:

$$\hat{h}(\omega_r) = \sum_{k=1}^{m_1} K_{m_1}\left(\frac{\omega_r - \tilde{\omega}_k}{\tilde{h}_0}\right) \cdot \tilde{h}(\tilde{\omega}_k),$$

where $K_{m_1}(t)$ is the local linear weight, similar to (2.6). The smoothed bandwidth is then used to estimate $f_Y(\omega_r)$.

Step 4. *Final spectral density estimate*: Obtain the estimate $\hat{f}_Y(\omega_r)$ in (2.6) by employing the bandwidths $\hat{h}(\omega_r)$, for $r = 1,\ldots,m_2$.

The bandwidth selector is not only easy to implement; the estimator is also stable in the peak regions where the bandwidths are smoothed by the local linear smoother in Step 3. This contrasts with the claim made by Fan and Gijbels ((1996), p. 242). The steps can be iterated using the estimate $\hat{f}_Y(\omega_r)$ in Step 4 as initial estimate in Step 1 and repeating Steps 2–4. The proposed spectral density estimation method can be easily applied to time series analysis with missing observations.

5. Simulations

We conduct a simulation study to compare the performance of the proposed method under a periodic missing pattern with the complete observation case. Note that for the complete observation case the proposed estimator is just the smoothed periodogram in Fan and Kreutzberger (1998), but for the incomplete observation case the latter estimator is not directly applicable.

For simplicity, we chose the popular Epanechikov kernel

$$K(x) = 0.75(1 - x^2)I(|x| \leq 1).$$

The data-driven local bandwidth selection rule in Section 4 was applied in the simulations. To compare the performance of different estimators in simulations, we use 'typical' dataset in the sense that, based on the dataset, the estimators have their median performance in terms of average squared deviation loss at grid points where the estimators are evaluated.

Consider the ARMA model

$$Y_t + a_1Y_{t-1} + \cdots + a_pY_{t-p} = \epsilon_t + b_1\epsilon_{t-1} + b_q\epsilon_{t-q}, \tag{5.1}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$. The following two examples were studied in Wahba (1980) and Fan and Kreutzberger (1998).

Example 1. The AR(3) model with $a_1 = -1.5$, $a_2 = 0.7$ and $a_3 = -0.1$ and the rest of the coefficients equal to zero.

Example 2. The MA(4) model with $b_1 = -0.3$, $b_2 = -0.6$, $b_3 = -0.3$ and $b_4 = 0.6$ and the rest of the coefficients equal to zero.

No missing and weekly missing patterns were considered for these examples. Here we specifically generated the missing pattern in each simulated series. We kept the simulated observations for five time points, then dropped the next two observations.

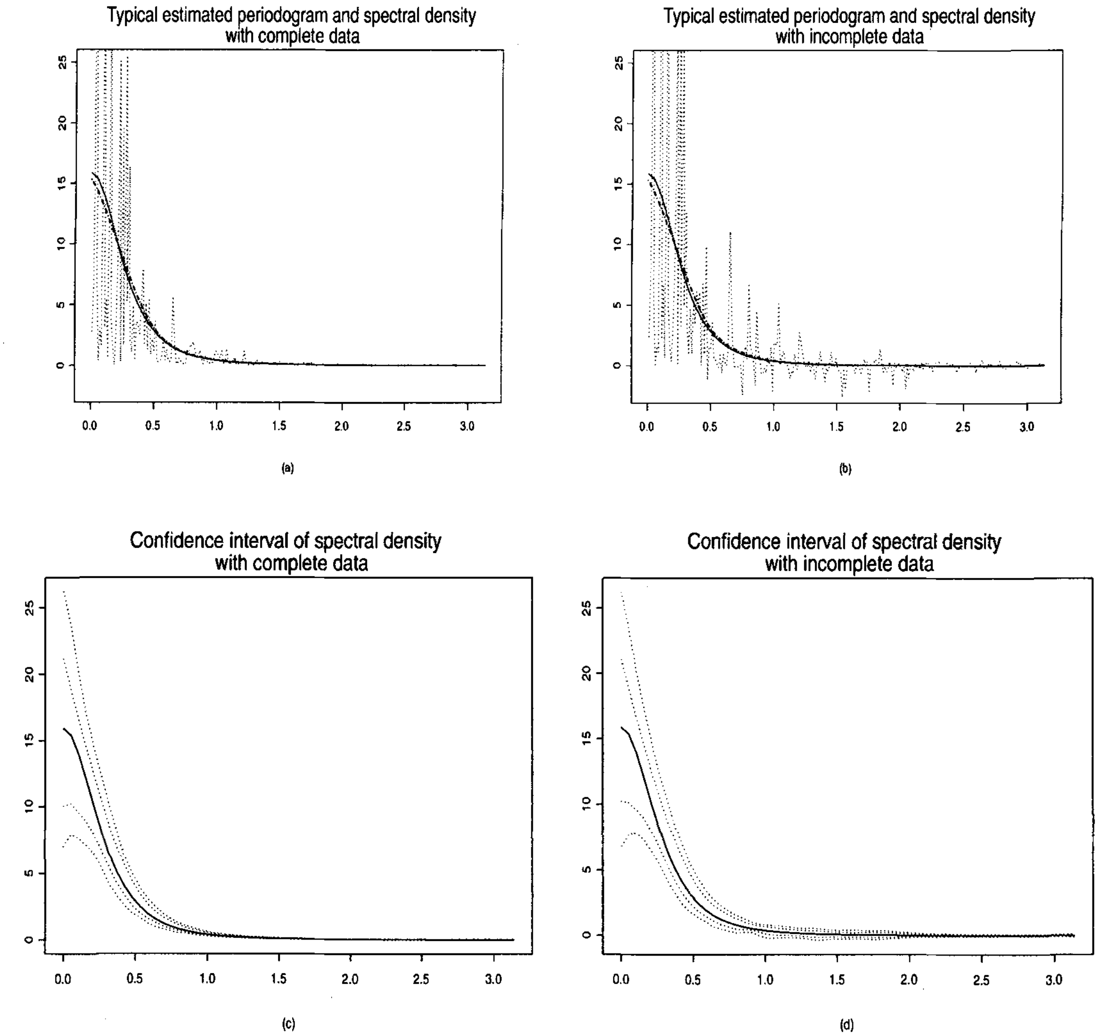

Fig. 1. Numerical results for Example 1.

Hence $\{g_t\}$ is a periodic function with period 7, and

$$g_t = \begin{cases} 1 & \text{if} \quad t = 1, \ldots, 5 \\ 0 & \text{if} \quad t = 6, 7. \end{cases}$$

We simulated 600 times with sample size $N = 250$. Two time series, a complete series and an incomplete series with specified missing pattern, were generated in each simulation.

Figures 1(a) and 1(b) give the 'typical' estimated periodograms (dotted lines) and spectral density functions (dashed lines) for the complete series and the incomplete series in Example 1 respectively. It is shown that the proposed estimator performs resaonably well for the incomplete series with missing observations (Fig. 1(b)) compared with the

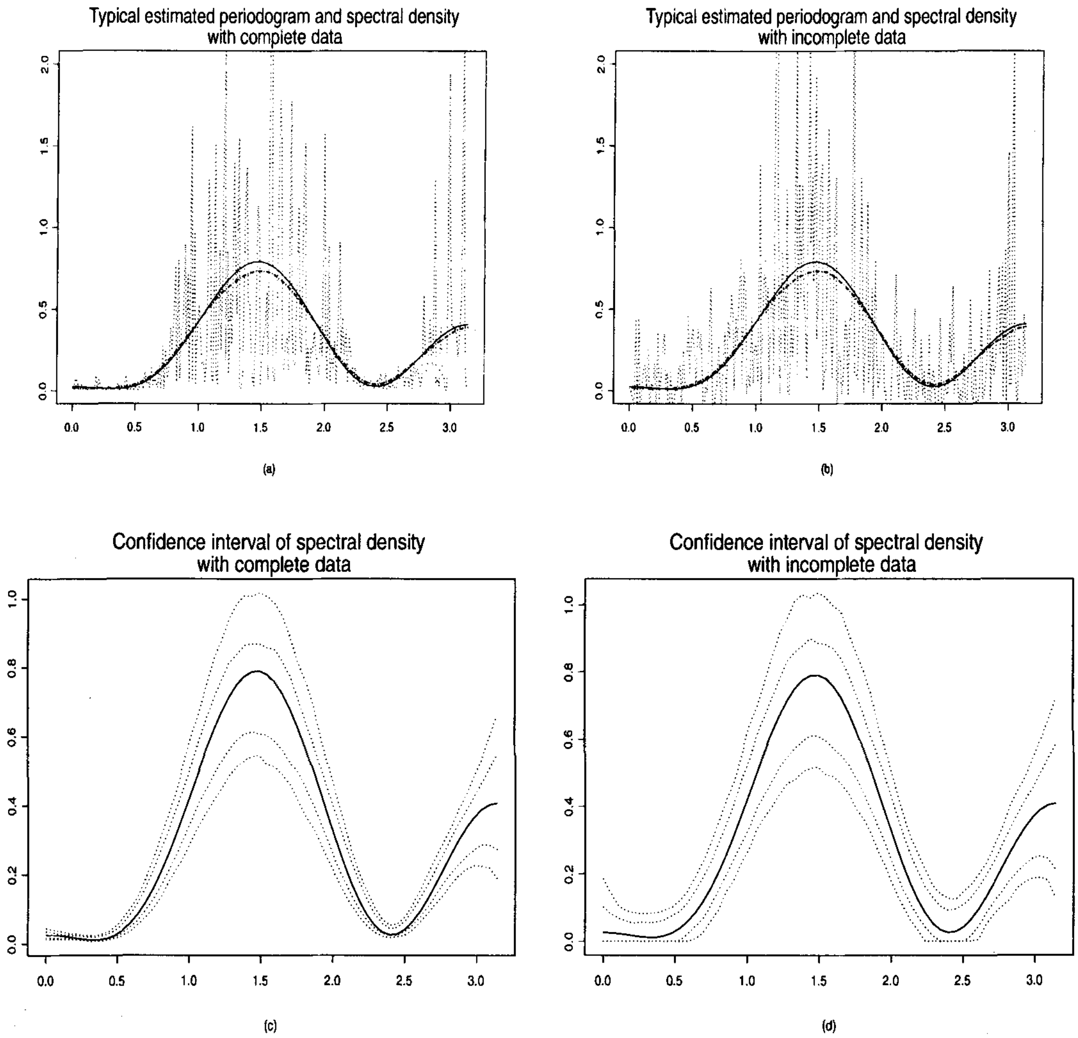

Fig. 2. Numerical results for Example 2.

complete series (Fig. 1(a)). The difference is insignificant. This demonstrates that the proposed estimator gives a good estimate under the missing observation situation. Figures 1(c) and 1(d) also plot the envelopes (dotted lines) of the true spectral density (solid line) constructed from the pointwise sample percentiles. These envelopes represent pointwise 2.5%, 12.5%, 87.5% and 97.5% sample percentiles of the spectral density estimates among 600 simulations. The estimated envelopes in Figs. 1(c) and 1(d) indicate that the proposed method gives good confidence bounds for the spectral density function when there are missing obervations. Therefore it is insensitive to missing observations. It also demonstrates that these confidence bounds provide a reliable tool in outlier detection, which is discussed further in the next section. Figure 2 plots the estimated spectral densities and envelopes for complete and incomplete series generated from Example 2. Results from Example 2 with a unimodal spectral density (Fig. 2) reconfirm the conclusion drawn from Example 1 with a J-shaped spectral density (Fig. 1).

It is shown for the above two examples that the proposed estimator performs reasonably well with about 30% missing observations in comparison to the case where complete observations available. To assess the effectiveness of the proposed estimator for a higher proportion of missing, we considered the following weekly missing pattern with a periodic function g_t such that

$$g_t = \begin{cases} 1 & \text{if} \quad t = 1, 3, 5, 7 \\ 0 & \text{if} \quad t = 2, 4, 6. \end{cases}$$

There are about 43% missing observations. For the previous two examples, the performance of the proposed estimator under this missing pattern is similar to that under the previous pattern, but with a slightly wider confidence band. The results will not be displayed here due to space constraint.

Example 3. *Continuation of Example* 2 *with random missing pattern.* In this example, we use the following random missing pattern to illustrate how to use the proposed

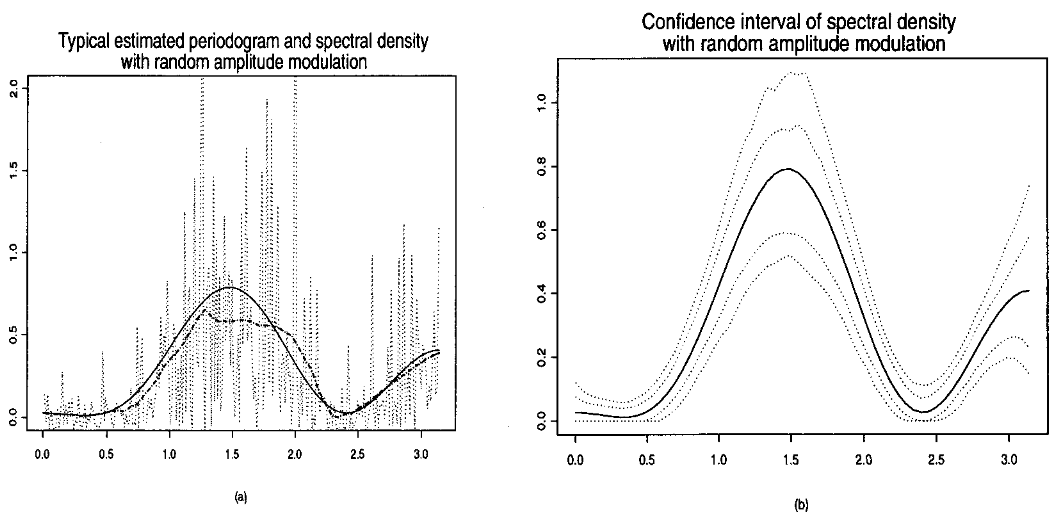

Fig. 3. Numerical results for Example 3.

estimation approach and to investigate if the method works in the setting:

$$P(g_t = 0 \mid g_{t-1} = 0) = 0.3, \qquad P(g_t = 0 \mid g_{t-1} = 1) = 0.2$$

with initial condition $g_0 = 1$ (i.e. $\{g_t\}$ is a Markov chain). For the MA(4) model in Example 2, we simulated 600 samples for $\{Y_t\}$ with size $N = 250$. For each sample, $\{g_t\}$ was sampled from the above Markov chain and $\{X_t\}$ was the observed series. Because no optimal bandwidth is available for the proposed estimator in this case, we use the one for the spectral density estimation on $\{Y_t\}$ in our simulations.

Figures 3(a) and 3(b) show the typical estimated curve and the pointwise 2.5%, 12.5%, 87.5% and 97.5% sample percentiles of the proposed estimators among 600 simulations for this example. Note that only $\{X_t\}$ was observed, direct use of the common smoothed periodogram for $\{X_t\}$ in Fan and Kreutzberger (1998) would result in a biased estimator of the spectral density of $\{Y_t\}$. Given $\{X_t\}$ and the transition probability matrix of $\{g_t\}$, we computed the proposed estimator. Figure 3 demonstrates that for the random missing pattern the proposed estimator truely captures the structure of the spectral density even though the bandwidths employed here are not optimal. Further study on the optimal choice of bandwidth in this case requires derivation of the asymptotic bias and the asymptotic variance of the estimator, which is worthy of further investigation.

6. Outlier detection

The existence of outliers may affect model identification, which is essential in time series analysis. See for example Abraham and Chuang (1989), Bruce and Martin (1989), Subba Rao (1989), and Tsay *et al.* (2000) among others. Our suggested method only requires a nonparametric estimation of spectral density which avoids the identification problem. Thus, it is expected that the diagnostic based on nonparametric spectral estimation be more reliable than common diagnostics based on parametric models.

The well-known masking and smearing phenomena post a difficult problem in the diagnostic tests for time series data. Earlier work in this area was presented in Chernick *et al.* (1982), Martin and Yohai (1986), and Bruce and Martin (1989). In spectral density estimation, outliers with a relatively large value compared with the scale of the innovation process will distort the estimator. These "outlying" points may not have a relatively large value compared with obervations' scale. We adopt the common leave-k-out diagnostic method to identify outliers in the spectral density estimation. The leave-k-out diagnostic has been widely used in regression analysis (see Cook and Weisberg (1982) and Atkinson (1985)). Bruce and Martin (1989) and Hui and Lee (1992) also studied the leave-one-out approach in the time series context. We expect that our diagnostic approach based on deletion and the proposed spectral density estimation identifies outliers (especially innovation outliers) without masking and smearing effects.

For illustration, we only consider the leave-one-out diagnostic method in outlier detection when there are missing observations. A multiple-deletion diagnostic can be developed in the same way, but we will not consider this problem since intensive computation would be required.

Diagnostic procedure for outlier detection

(1) *Confidence bound.* Given a significance level α (1% or 5%, say), construct a confidence bound for the spectral density with all available observations based on the

estimators in (2.6) and the normal approximation in Theorem 3.1:

$$\hat{f}_Y(\omega) - \widehat{Bias}(\hat{h}_{\mathrm{opt}}(\omega), \omega) \pm Z_{1-\alpha/2} \cdot \left(\frac{v_0(K)\pi[\hat{f}_Y^2(\omega) + \hat{\tau}_Y^2(\omega)]}{N\hat{h}_{\mathrm{opt}}(\omega)} \right)^{1/2},$$

where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of standard normal distribution.

(2) *Spectral density estmation with data deletion.* Apply the leave-one-out approach to estimate the spectral density. Evaluate the spectral density curves in (2.6) with the i-th observation deleted where observation i is regarded as periodic missing with period $\theta = n$.

(3) *Outlier identification.* The i-th observation is identified as an outlier if the estimated spectral density curve with i-th observation deleted exceeds the confidence bounds in (1).

Example 4. Consider the yearly differenced RESEX series consisting of 77 observations with two consecutive "outliers" at $t = 71$ and $t = 72$. The original RESEX dataset is given in Martin *et al.* (1983), and consists of Bell Canada inward movement of residential telephone extensions in a fixed geographic area from January 1966 to May 1973. As pointed out by Hui and Lee (1992), several common diagnostic approaches suffer from masking and smearing phenomena, which motivated Jiang *et al.* (1999) to use the robust L_1-norm fit of the dataset. Here we use the proposed diagnostic procedure to identify outliers in the data.

The data-driven bandwidth selection rule in Section 4 and the proposed diagnostic test were applied in this example. Figure 4(a) gives the estimated periodogram (solid line) and spectral density (dotted line), and Fig. 4(b) reports the 99% confidence bounds (dotted-dash lines) and the corresponding estimated spectral density curves with one observation in the dataset deleted as missing. It is shown that only the estimated curves with observations 71 and 72 deleted respectively stretch out of the confidence bounds. The two observations $t = 71$ and $t = 72$ are correctly identified as outliers.

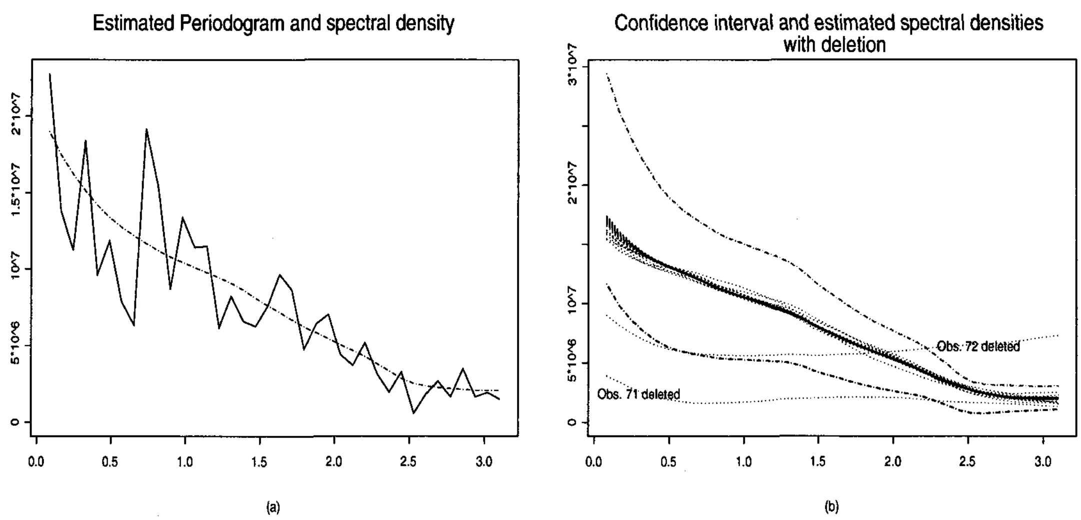

Fig. 4. Numerical results for Example 4.

The illustrative example also shows that the diagnostic procedure avoids masking and smearing phenomena.

7. Conclusion

We have considered the spectral density estimation for a deterministic amplitude modulated pattern, especially for Parzen's periodic modulation mechanics. The performance of the estimator applied to an incomplete series is demonstrated in a simulation study. The proposed method also works for a random modulation pattern such that $\{g_t\}$ is a stationary process with mean $E[g_t]$ and is independent of $\{Y_t\}$. Under this situation, $\{X_t\}$ is stationary with covariance function $R_X(v) = R_Y(v)R_g(v)$, where $R_g(v) = E[g_t g_{t+v}]$. Then one can construct the estimator for spectral density based on smoothing the generalized periodograms, where $R_Y(v) = R_X(v)/R_g(v)$ can be estimated by its sample average.

The proposed outlier detection approach can be applied to other diagnostic tests. For example, one may also identify influential cases by employing the estimated spectral curve in (2.6) and the methods mentioned in Subba Rao (1989) and Hui and Lee (1992).

Acknowledgements

The authors gratefully acknowledge helpful comments and suggestions from Professor Jianqing Fan.

Appendix

Since the main result will be derived from the proof for Theorem 3.1, we first present the proof of Theorem 3.1 and then discuss the proof of Proposition 2.1. To facilitate the exposition, the following notations from Parzen (1963) are adopted:

$$\lambda_m = 2\pi m/\theta, \qquad \text{for} \quad m = 1, \ldots, \Theta$$

$$h(s,t) = \frac{g_s g_t}{R_g(s-t)} = \sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n} \exp[i(s\lambda_m + t\lambda_n)].$$

PROOF OF THEOREM 3.1. We only give the proof for interior points. For boundary points $\omega_n^* = ch$, the proof is basically similar to that for the interior points. Note that

$$\hat{R}_Y(v) = \frac{1}{n} \sum_{t=1}^{n-v} g_t g_{t+v} Y_t Y_{t+v} / R_g(v).$$

It follows by the definition of $GI_Y^{(n)}(\omega_k)$ that

$$\text{(A.1)} \qquad GI_Y^{(n)}(\omega_k) = \frac{1}{2\pi} \sum_{|v|<n} \hat{R}_Y(v) \exp(-iv\omega_k)$$

$$= \frac{1}{2\pi n} \sum_{s=1}^{n} \sum_{t=1}^{n} h(s,t) \exp[-i(s-t)\omega_k] Y_s Y_t.$$

Denote the discrete Fourier transforms of $\{Y_t\}$ by $\mathcal{J}_Y(\omega) = \frac{1}{\sqrt{2\pi n}}\sum_{s=1}^{n}\exp(-is\omega)Y_s$. Then $GI_Y^{(n)}(\omega)$ can be rewritten as

$$\begin{aligned}
\text{(A.2)} \qquad GI_Y^{(n)}(\omega) &= \frac{1}{2\pi n}\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n} \\
&\qquad \times \sum_{s,t=1}^{n} \exp[i(s\lambda_m + t\lambda_n)]\exp[-i(s-t)\omega]Y_sY_t \\
&= \sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\frac{1}{\sqrt{2\pi n}}\sum_{s=1}^{n}\exp[-is(\omega-\lambda_m)]Y_s \\
&\quad \times \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}\exp[it(\omega+\lambda_n)]Y_t \\
&\equiv \sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\mathcal{J}_Y(\omega-\lambda_m)\overline{\mathcal{J}_Y(\omega+\lambda_n)}.
\end{aligned}$$

From the proof of Theorem 10.3.1 in Brockwell and Davis (1991), we have

$$\text{(A.3)} \qquad \mathcal{J}_Y(\omega) = \Psi(e^{-i\omega})\mathcal{J}_Z(\omega) + Y_n(\omega),$$

where $Y_n(\omega) = \frac{1}{\sqrt{2\pi n}}\sum_{j=-\infty}^{\infty}\psi_j e^{-ij\omega}U_{nj}$ such that $E|Y_n(\omega)|^4 = O(n^{-2})$ and $U_{nj} = \sum_{s=1-j}^{n-j} Z_s e^{-is\omega} - \sum_{s=1}^{n} Z_s e^{-is\omega}$. Combining (A.2) and (A.3) gives

$$\begin{aligned}
GI_Y^{(n)}(\omega) &= \sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\Psi(e^{-i(\omega-\lambda_m)})\mathcal{J}_Z(\omega-\lambda_m)\overline{\Psi(e^{-i(\omega+\lambda_n)})\mathcal{J}_Z(\omega+\lambda_n)} \\
&\quad + \Bigg\{\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}[\Psi(e^{-i(\omega-\lambda_m)})\mathcal{J}_Z(\omega-\lambda_m)\overline{Y_n(\omega+\lambda_n)} \\
&\qquad\qquad + Y_n(\omega-\lambda_m)\overline{\Psi(e^{-i(\omega+\lambda_n)})\mathcal{J}_Z(\omega+\lambda_n)} \\
&\qquad\qquad + Y_n(\omega-\lambda_m)\overline{Y_n(\omega+\lambda_n)}]\Bigg\} \\
&\equiv A_{n1}(\omega) + A_{n2}(\omega).
\end{aligned}$$

Therefore

$$\begin{aligned}
\text{(A.4)} \qquad \hat{f}_Y(\omega) &= \sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right)A_{n1}(\omega_k) + \sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right)A_{n2}(\omega_k) \\
&\equiv M_n(\omega) + R_n(\omega).
\end{aligned}$$

Observe that $\max_{\omega\in[0,\pi]} E|Y_n(\omega)|^4 = O(n^{-2})$, so we have $A_{n2}(\omega) = O_p(\frac{1}{\sqrt{n}})$ uniformly in $\omega \in [0,\pi]$. Hence $R_n(\omega) = O_p(\frac{1}{\sqrt{n}})$, which is an asymptotically negligible term, and suffices to derive the asymptotic distribution of $M_n(\omega)$.

Note that

$$
\begin{aligned}
A_{n1}(\omega) &= \sum_{s,t=1}^{n} \frac{1}{2\pi n} \\
&\quad \times \left[\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n} e^{i(s\lambda_m + t\lambda_n)} \Psi(e^{-i(\omega-\lambda_m)}) \overline{\Psi(e^{-i(\omega+\lambda_n)})} e^{-i(s-t)\omega} \right] Z_s Z_t \\
&\equiv \sum_{s,t=1}^{n} r_\omega(s,t) Z_s Z_t.
\end{aligned}
$$

Observing $\overline{H_{m,n}} = H_{-n,-m}$, we have $\overline{r_\omega(s,t)} = r_\omega(t,s)$. Then $A_{n1}(\omega)$ is real. Therefore by (A.4),

$$
\begin{aligned}
M_n(\omega) &= \sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) \sum_{s,t=1}^{n} r_{\omega_k}(s,t) Z_s Z_t \\
&= \sum_{s=1}^{n} \left[\sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) r_{\omega_k}(s,s) \right] Z_s^2 \\
&\quad + \sum_{1\le s\ne t\le n} \left[\sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) r_{\omega_k}(s,t) \right] Z_s Z_t \\
&\equiv B_{n1}(\omega) + B_{n2}(\omega).
\end{aligned} \tag{A.5}
$$

We show that $B_{n1}(\omega)$ contributes to the bias, and $B_{n2}(\omega)$ to the variance of $M_n(\omega)$.

Direct calculation gives

$$
\begin{aligned}
E[B_{n1}(\omega)] &= \sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) f_Y(\omega_k) \\
&= f_Y(\omega) + \frac{1}{2} h^2 f_Y''(\omega)\mu_2(K) + o(h^2),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}[B_{n1}(\omega)] &= \sum_{s=1}^{n} \left[\sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right) r_{\omega_k}(s,s) \right]^2 2\sigma^4 \\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

Then

$$
B_{n1}(\omega) = f_Y(\omega) + \frac{1}{2} h^2 f_Y''(\omega)\mu_2(K) + o(h^2) + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.6}
$$

It remains to show that

$$
W(n) \equiv \sqrt{Nh} B_{n2}(\omega) \xrightarrow{\mathcal{D}} \mathcal{N}(0, v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)]). \tag{A.7}
$$

Note that

$$W(n) = \sum_{1\le s\ne t\le n} \left[\sum_{k=1}^{N} \sqrt{Nh} K_N\left(\frac{\omega-\omega_k}{h}\right) r_{\omega_k}(s,t)\right] Z_s Z_t$$
$$\equiv \sum_{1\le s\ne t\le n} b_{s,t}(\omega) Z_s Z_t.$$

Observe that $\overline{b_{s,t}(\omega)} = b_{t,s}(\omega)$ for $s \ne t$, hence it follows that $a_{s,t}(\omega) \equiv [b_{s,t}(\omega)+b_{t,s}(\omega)]/2$ is real. Then

$$\text{(A.8)} \qquad W(n) = \sum_{1\le s<t\le n} [b_{s,t}(\omega) + b_{t,s}(\omega)] Z_s Z_t = \sum_{s,t=1}^{n} a_{s,t}(\omega) Z_s Z_t,$$

where $a_{s,s}(\omega) = 0$ for $s = 1,\ldots,n$. Let $A(\omega) = (a_{s,t}(\omega))$ be $n\times n$ matrix, then $A(\omega)$ is symmetric with vanishing diagonal elements. Applying Theorem 5.2 in de Jong (1987) to (A.8), one gets (A.7). This completes the proof of the theorem. De Jong's Theorem is stated in the following.

THEOREM A.1. *Let $Q(n) = \sum_{1\le i\ne j\le n} a_{ij} Z_i Z_j$ be a quadratic form in independent random variables Z_i ($EZ_i = 0, EZ_i^2 = 1$), with $\nu_1,\ldots,\nu_n$ the eigenvalues of the symmetric matrix (a_{ij}), with vanishing diagonal elements: $a_{ii} = 0$ for all i. Suppose there exists a sequence of real numbers $c(n)$ such that (let $\sigma(n)^2$ be the variance of $Q(n)$):*

1) *$c^4(n)\sigma(n)^{-2}\max_{1\le i\le n}\sum_{1\le j\le n} a_{i,j}^2 \to 0$, as $n\to\infty$; and*

2) *$\max_{1\le i\le n} E[Z_i^2 I(|Z_i| > c(n))] \to 0$, as $n\to\infty$.*

3) *If the eigenvalues of the matrix (a_{ij}) are negligible: $\sigma(n)^{-2}\max_{1\le i\le n}\nu_i^2 \to 0$, as $n\to\infty$ then*

$$\sigma(n)^{-1} Q(n) \xrightarrow{d} N(0,1), \qquad n\to\infty.$$

To apply the theorem, we need to check the following conditions:

(1) $A(\omega)$ is symmetric and has vanishing diagonal elements;

(2) $\sigma(n)^2 \equiv \mathrm{Var}[W(n)] = v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)] + o(1)$;

(3) there exists a sequence of real numbers $c(n)\to\infty$ such that

$$c^4(n)\sigma(n)^{-2}\max_{1\le i\le n}\sum_{1\le j\le n} a_{i,j}^2(\omega) \to 0, \qquad \text{as} \quad n\to\infty;$$

(4) $\max_{1\le i\le n} E[Z_i^2 I(|Z_i| > c(n))] \to 0$, as $n\to\infty$; and

(5) $\sigma(n)^{-2}\max_{1\le i\le n}\mu_i^2 \to 0$, as $n\to\infty$, where μ_i's are the eigenvalues of the matrix $A(\omega)$.

Conditions (1) and (4) are obvious from the definition of $W(n)$ and the normality of Z_i if there exists a sequence of real numbers $c(n)\to\infty$ in condition (3). Note that

$$\max_{1\le i\le n}\mu_i^2 \le \sum_{i=1}^{n}\mu_i^2 = n^{-1}\,\mathrm{tr}(A^2(\omega)) = n^{-1}\sum_{s,t=1}^{n} a_{s,t}^2(\omega).$$

It follows that condition (5) holds if $n^{-1}\sum_{s,t=1}^{n} a_{s,t}^2(\omega) \to 0$. Let

$$\Psi_{m,n}(\lambda) = \Psi(e^{-i(\lambda-\lambda_m)})\overline{\Psi(e^{-i(\lambda+\lambda_n)})}$$

and

$$\Phi_{m,n}(\lambda; s,t) = e^{-is(\lambda-\lambda_m)}\overline{e^{-it(\lambda+\lambda_n)}} + e^{-it(\lambda-\lambda_m)}\overline{e^{-is(\lambda+\lambda_n)}}.$$

Then

$$r_\lambda(s,t) = \frac{1}{2n\pi}\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\Psi_{m,n}(\lambda)e^{-is(\lambda-\lambda_m)}\overline{e^{-it(\lambda+\lambda_n)}}.$$

From the definition of $a_{s,t}(\omega)$, direct computation leads to

$$\begin{aligned}
a_{s,t}(\omega) &= \frac{\sqrt{Nh}}{2}\sum_{k=1}^{n} K_N\left(\frac{\omega-\omega_k}{h}\right)[r_{\omega_k}(s,t) + \overline{r_{\omega_k}(s,t)}] \\
&= \frac{\sigma^{-2}}{2}\sqrt{\frac{h}{2n}}\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\frac{\sigma^2}{2\pi}\sum_{k=1}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right)\Psi_{m,n}(\omega_k)\Phi_{m,n}(\omega_k; s,t) \\
&= \frac{\sigma^{-2}}{2}\sqrt{\frac{h}{2n}}\sum_{m,n=-\Theta}^{\Theta} e_m e_n H_{m,n}\frac{\sigma^2}{2\pi}\Psi_{m,n}(\omega)\Phi_{m,n}(\omega; s,t)(1+O(h^2)),
\end{aligned}$$

uniformly for s,t with $s \neq t$. Then $\sum_{t=1}^{n} a_{s,t}^2(\omega) = O(h)$ uniformly for $s = 1,\ldots,n$. Hence

$$\max_{1\le s\le n}\sum_{t=1}^{n} a_{s,t}^2(\omega) \to 0 \quad \text{and} \quad n^{-1}\sum_{s,t=1}^{n} a_{s,t}^2(\omega) \to 0.$$

Conditions (3) and (5) hold if we take $h \to 0$ and $c(n) \to \infty$. Condition (2) holds from the results of (A.4)–(A.7) and Lemma A.1 below. □

LEMMA A.1. *Under assumptions* (i)–(v), *we have for each* $\omega \in (0,\pi)$

$$\mathrm{Var}[M_n(\omega)] = \frac{1}{Nh}v_0(K)\pi[f_Y^2(\omega) + \tau_Y^2(\omega)](1+o(1)).$$

PROOF. Note that

$$A_{n1}(\omega) = \sum_{s,t=1}^{n} r_\omega(s,t)Z_sZ_t.$$

Following the same argument as that given in Parzen ((1963), equation 3.29), we have

$$\mathrm{Var}[A_{n1}(\omega)] = f_Y^2(\omega) + \tau_Y^2(\omega) + O\left(\frac{1}{n}\right) \tag{A.9}$$

and

$$\mathrm{Cov}(A_{n1}(\omega_k), A_{n1}(\omega_{k'})) = O\left(\frac{1}{n}\right), \tag{A.10}$$

uniformly in k, k'. Then

$$\begin{aligned}\text{(A.11)} \qquad \mathrm{Var}[M_n(\omega)] &= \sum_{k=1}^{N} K_N^2\left(\frac{\omega-\omega_k}{h}\right)\mathrm{Var}[A_{n1}(\omega_k)] \\ &\quad + \sum_{k\neq k'}^{N} K_N\left(\frac{\omega-\omega_k}{h}\right)K_N\left(\frac{\omega-\omega_{k'}}{h}\right) \\ &\quad \times \mathrm{Cov}(A_{n1}(\omega_k), A_{n2}(\omega_{k'})) \\ &\equiv D_{n1}+D_{n2}.\end{aligned}$$

By (A.9), (A.10) and (A.11), simple algebra leads to

$$D_{n1} = \frac{1}{Nh}\pi v_0(K)[f_Y^2(\omega)+\tau_Y^2(\omega)](1+o(1))$$

and

$$D_{n2} = O\left(\frac{1}{n}\right),$$

which completes the proof the lemma. □

PROOF OF PROPOSITION 2.1. The proof is similar to the proof of Lemma A.1, and is shown by employing the arguments for (3.2), (3.3) and (3.29) in Parzen (1963). □

REFERENCES

Abraham, B. and Chuang, A. (1989). Outlier detection and time series modeling, *Technometrics*, **31**, 241–248.

Atkinson, A. C. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press, Oxford.

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

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

Bruce, A. G. and Martin, R. D. (1989). Leave-k-out diagnostics for time series, *Journal of the Royal Statistical Society, Series B*, **51**, 363–424.

Chernick, M. R., Downing, D. J. and Pike, D. H. (1982). Detecting outliers in time series data, *Journal of the American Statistical Association*, **77**, 743–747.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall, New York.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms, *Probability Theory and Related Fields*, **75**, 261–277.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation, *Journal of the Royal Statistical Society, Series B*, **57**, 371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, New York.

Fan, J., Jiang, J., Zhang, C. M. and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics, *Statistica Sinica*, **13**, 965–992.

Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation, *Scandinavian Journal of Statistics*, **25**, 359–369.

Hui, Y. V. and Lee, A. H. (1992). Assessing influence in time series, *Communications in Statistics—Theory and Methods*, **21**, 3463–3478.

Jiang, J., Hui, Y. V. and Zheng, Z. (1999). Robust goodness-of-fit tests for AR(p) models based on L_1-norm fitting, *Science in China*, **42**, 337–346.

Jones, R. H. (1962). Spectral analysis with regularly missed observations, *The Annals of Mathematical Statistics*, **32**, 455–461.

Kleiner, B., Martin, R. D. and Thomson, D. J. (1979). Robust estimation of power spectra (with discussion), *Journal of the Royal Statistical Society, Series B*, **41**, 313–351.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995*a*). Rate of convergence for logspline spectral density estimation, *Journal of Time Series Analysis*, **16**, 389–401.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995*b*). Logspline estimation of a possibly mixed spectral distribution, *Journal of Time Series Analysis*, **16**, 359–388.

Martin, R. D., Samarov, A. and Vandaele, W. (1983). Robust methods for ARIMA models, *Applied Time Series Analysis of Economic Data* (ed. A. Zellner), 153–169, US Bureau of Census, Washington D.C.

Martin, R. D. and Yohai, V. J. (1986). Influence functionals for time series, *The Annals of Statistics*, **14**, 781–818.

Müller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths, *Biometrics*, **50**, 61–76.

Parzen, E. (1961). Mathematical considerations in the estimation of spectra, *Technometrics*, **3**, 167–190.

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation, *Sankhyā, Series A*, **25**, 383–392.

Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood, *Journal of the American Statistical Association*, **89**, 600–610.

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

Ruppert, D. (1997). Empirical bias bandwidths for local polynomial nonparametric regression and density estimation, *Journal of the American Statistical Association*, **92**, 1049–1062.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association*, **90**, 1257–1270.

Subba Rao, T. (1989). Discussion of the paper by Bruce and Martin, *Journal of the Royal Statistical Society, Series B*, **41**, 313–351.

Tsay, R. S., Peña, D. and Pankratz, A. E. (2000). Outliers in multivariate time series, *Biometrika*, **87**, 789–804.

Wahba, G. (1980). Automatic smoothing of the log periodogram, *Journal of the American Statistical Association*, **75**, 122–132.