

Automatic Local Smoothing for Spectral Density Estimation

Jianqing Fan

Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Eva Kreutzberger

Department of Mathematics
University of Kaiserslautern
67653 Kaiserslautern, Germany

Abstract

This article uses local polynomial techniques to fit Whittle's likelihood for spectral density estimation. Asymptotic sampling properties of the proposed estimators are derived, and adaptation of the proposed estimator to the boundary effect is noted. We show that the Whittle likelihood based estimator has advantages over the least-squares based log-periodogram. The bandwidth for the Whittle likelihood-based method is chosen by a simple adjustment of a bandwidth selector proposed in Fan and Gijbels (1995). The effectiveness of the proposed procedure is demonstrated by a few simulated and real numerical examples. Our simulation results support the asymptotic theory that the likelihood based spectral density and log-spectral density estimators are the most appealing among their peers.

KEY WORDS: Bandwidth selection, local polynomial fit, periodogram, spectral density estimation, Whittle likelihood.

1 Introduction

Spectral density estimation is useful for studying stationary time series, including periodicity. The raw material used to construct a spectral density estimate is periodogram. Let $X_t, t = 0, \pm 1, \pm 2, \dots$ denote a zero mean stationary process with the autocovariance function

$$\gamma(u) = E(X_s X_{s+u}), \quad u = 0, \pm 1, \pm 2, \dots$$

Then, the periodogram for an observed time series X_1, \dots, X_n is given by

$$I^{(n)}(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n e^{-i\lambda t} X_t \right|^2, \quad \lambda \in [0, \pi].$$

It is well known that the periodogram is approximately an unbiased estimator of the spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma(u) \exp(-iu\lambda), \quad \lambda \in [0, \pi].$$

But, it is not a consistent estimator of $f(\lambda)$. See for example Brillinger (1981), Priestley (1981) and Brockwell and Davis (1991).

Consistent estimators of $f(\lambda)$ can be obtained by smoothing the periodogram. Let $\lambda_k = 2\pi k/n$ be a Fourier frequency. There are three possible ways to obtain a consistent estimate. The first approach is to directly smooth on the data $\{(\lambda_k, I^{(n)}(\lambda_k)), k = 1, \dots, [(n-1)/2]\}$

via a weighted local average. Most of the earlier literature used this approach. See Chapter 5 of Brillinger (1981) and references therein. The second approach is to smooth over the log-periodogram $\{(\lambda_k, \log I^{(n)}(\lambda_k))\}$ via a least-squares method. See for example Wahba (1980) who used a smoothing spline minimizing penalized square errors. The third approach is based on the Whittle (1962) likelihood of the periodogram. See for example, Pawitan and O'Sullivan (1994) and Kooperberg, Stone and Truong (1995a, b) where spline techniques are employed. As in all nonparametric smoothing problems, an important issue is to choose appropriate smoothing parameters. Extensive efforts have been made in the literature. See for example Swanepoel and van Wyk (1986), Beltrão and Bloomfield (1987), Hurvich and Beltrão (1990), Franke and Härdle (1992) and Politis and Romano (1992). Estimation of locations of peaks was studied by Newton and Pagano (1983) and Müller and Prewitt (1992).

There is also a large literature on spectral density estimation based on ARMA or other parametric models. See for example §10.6 of Blockwell and Davis (1991). The basic idea is to fit an ARMA model with orders selected adaptively by some criterion such as the Bayesian Information Criterion. These methods can be very useful for many applications. However, the computation cost can be expensive and not all spectral densities can efficiently be approximated by ARMA models.

Among the three possible smoothing schemes mentioned above, which one is more appealing? This problem has not been convincingly answered yet. It is also not easy to compare the above approaches since they are based on different smoothing procedures. In this paper, we study the theoretical and numerical properties of the above three approaches to spectral density estimation using local polynomial fitting as a common smoother. This smoothing method is *asymptotically* equivalent to kernel method (Müller, 1987) except at boundary regions, when designs are equispaced. Recent developments on the local polynomial approach can be found in Fan (1992), Hastie and Loader (1993), Ruppert and Wand (1994), Fan, Heckman and Wand (1995), Jones (1997), among others.

We show that the Whittle likelihood based estimator has a smaller asymptotic variance than the least-squares smoothed log-periodogram, while their asymptotic biases are the same. We also demonstrate that the likelihood based estimator has a smaller bias than the smoothed periodogram at regions where the log-spectral density is convex (the biases are not comparable at concave regions), while maintaining the same asymptotic variance as the smoothed periodogram. Hence, the Whittle likelihood based method is recommended.

The bandwidth of the Whittle likelihood based local polynomial fitting is selected via a simple adjustment of a pre-asymptotic substitution method developed in Fan and Gijbels (1995). The effectiveness of this method is convincingly demonstrated via intensive simulations. For comparisons with least-squares based approaches, we also give data-driven bandwidth selectors, based on the methods of Fan and Gijbels (1995), for the least-squares smoothed periodogram

and log-periodogram. Three data-driven methods for spectral density estimation are then compared via intensive simulations. It is shown in Section 3 that for the simulation models considered the likelihood based method are the best among three procedures considered. This is consistent with our asymptotic theory.

We begin by describing the Whittle likelihood based spectral density estimation in Section 2. Finite sample comparisons of three spectral density estimators are presented in Section 3. The asymptotic properties of the likelihood based estimator is summarized in Section 4 and the proofs are given in Section 5.

2 Estimation of spectral density

Let $\lambda_k = 2\pi k/n$ ($k = 0, 1, \dots, [(n-1)/2]$) be Fourier frequencies and let $N = [(n-1)/2]$. Given a stationary linear process $\{X_t\}$, it is known (see, for example, Theorem 10.3.2 of Brockwell and Davis, 1991) that $\{I^{(n)}(\lambda_k)\}$ are asymptotically exponentially distributed with mean $f(\lambda_k)$ and that they are approximately independent. That is, with $R_{n,k}$ denoting an asymptotically negligible term,

$$I^{(n)}(\lambda_k) = f(\lambda_k) \cdot V_k + R_{n,k}, \quad (2.1)$$

where the random variables V_k 's are independent having the standard exponential distribution for $k = 1, \dots, N$. Moreover, V_0 and $V_{[n/2]}$ (if n is even) have a χ_1^2 -distribution. For convenience, we will only consider the periodogram $I^{(n)}(\lambda_k)$ at frequencies λ_k , $k = 1, \dots, N$.

Thus the logarithm of the periodogram exhibits the following regression model:

$$\begin{cases} Y_k = \log I^{(n)}(\lambda_k) = m(\lambda_k) + \varepsilon_k + r_k, & k = 1, \dots, N, \\ \varepsilon_k = \log(V_k) \text{ has a density } f_\varepsilon(x) = \exp(-\exp(x) + x) \end{cases}, \quad (2.2)$$

where $m(\lambda_k) = \log f(\lambda_k)$ and $r_k = \log[1 + R_{n,k}/\{f(\lambda_k)V_k\}]$ is an asymptotically negligible term. It is known that $E(\varepsilon_k) = C_0 = -.57721$, an Euler constant, and $\text{var}(\varepsilon_k) = \pi^2/6$. See Davis and Jones (1968).

Regarding (2.2) as a nonparametric regression model and applying the local linear smoother, we obtain the smoothed log-periodogram

$$\hat{m}_{\text{LS}}(\lambda) = \hat{a} = \sum_{k=1}^N w_k(\lambda)(Y_k - C_0),$$

where $w_k(\lambda)$ is the local linear weight. See for example Fan (1992). We can also apply a higher order local polynomial regression (Stone 1977, and Cleveland 1979) to estimate the log-spectral density m , but decide not to use it for the sake of simplicity.

It can be shown under Conditions (i) – (iv) in Section 4 that asymptotically

$$\hat{m}_{\text{LS}}(\lambda) - m(\lambda) \sim N\left(h^2 m''(\lambda) \mu_2(K)/2, (\pi^2/6) \nu_0(K) \pi/(nh)\right), \quad (2.3)$$

where $\mu_2(K) = \int u^2 K(u) du$ and $\nu_0(K) = \int K^2(u) du$. Thus, the asymptotically optimal bandwidth, which minimizes the integrated asymptotic squared bias and variance, is given by

$$h_{\text{LS, OPT}} = \left[\frac{\nu_0(K)(\pi^2/6)\pi}{\mu_2^2(K) \int_0^\pi \{m''(\lambda)\}^2 d\lambda} \right]^{1/5} n^{-1/5}. \quad (2.4)$$

In practical implementations, the optimal bandwidth has to be estimated. Since model (2.2) is a special case of the general nonparametric regression model, we can apply methods developed in the nonparametric regression literature. In particular, the constant bandwidth selector proposed in Fan and Gijbels (1995) will be used for selecting the bandwidth of \hat{m}_{LS} .

The least-squares based estimator \hat{m}_{LS} is not efficient, since the distribution of ε_k is not normal. Its efficiency can be improved by using the local maximum likelihood method.

For each given λ , we form the weighted log-likelihood as follows:

$$\mathcal{L}(a, b) = \sum_{k=1}^N [-\exp\{Y_k - a - b(\lambda_k - \lambda)\} + Y_k - a - b(\lambda_k - \lambda)] K_h(\lambda_k - \lambda), \quad (2.5)$$

where $K_h(\cdot) = K(\cdot/h)/h$. Let \hat{a} and \hat{b} be the maximizers of (2.5). The proposed local likelihood estimator for $m(\lambda)$ is $\hat{m}_{\text{LK}}(\lambda) = \hat{a}$.

The weighted log-likelihood (2.5) is similar to the Whittle (1962) likelihood based on the exponential model (2.1) with $R_{n,k} = 0$ except that the kernel weight is introduced to localize the approximation $m(x) \approx m(\lambda) + m'(\lambda)(x - \lambda)$. It is a strictly concave function so that there is a unique minimizer.

It will be shown in Section 4 that $\hat{m}_{\text{LK}}(\lambda)$ has the following asymptotic bias and variance:

$$\hat{m}_{\text{LK}}(\lambda) - m(\lambda) \sim N\left(h^2 m''(\lambda) \mu_2(K)/2, \pi \nu_0(K)/(nh)\right).$$

Comparing with (2.3), we note that the asymptotic variance of \hat{m}_{LK} is a factor $\pi^2/6 = 1.645$ smaller than that of the least-squares estimator $\hat{m}_{\text{LS}}(\lambda)$. Defining the optimal bandwidth in a similar fashion to (2.4), we find that the optimal bandwidth is

$$h_{\text{LK, OPT}} = (6/\pi^2)^{1/5} h_{\text{LS, OPT}} = 0.9053 h_{\text{LS, OPT}}.$$

Therefore, $\hat{m}_{\text{LK}}(\cdot)$ can use the optimal bandwidth for estimator $\hat{m}_{\text{LS}}(\cdot)$ adjusted by a factor of 0.9053. In the implementation below, the bandwidth for \hat{m}_{LK} is chosen as

$$\hat{h}_{\text{LK, OPT}} = (6/\pi^2)^{1/5} \hat{h}_{\text{LS, OPT}},$$

where $\hat{h}_{\text{LS, OPT}}$ is the constant bandwidth selector in Fan and Gijbels (1995) for the least-squares nonparametric problem (2.2).

Maximization of (2.5) can be done via the Newton-Raphson algorithm or the Fisher scoring method. The least-squares estimator can be used as an initial value. Since the least-squares

estimator possesses the right rate of convergence, the one-step iteration of the Newton-Raphson algorithm will produce an estimator that has the same asymptotic efficiency as \hat{m}_{LK} . A similar result was proved by Fan and Chen (1997) in the generalized linear model setting, which includes model (2.2) with $r_k = 0$. Thus, if computation cost is of concern, one can use the one-step estimator instead of \hat{m}_{LK} to improve the efficiency of \hat{m}_{LS} .

To illustrate the performance of \hat{m}_{LK} with the data-driven bandwidth, we use the ARMA model

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

to generate testing examples, where $\varepsilon_t \sim N(0, \sigma^2)$. The model has the spectral density

$$\frac{\sigma^2 |1 + b_1 \exp(-i\lambda) + \cdots + b_q \exp(-iq\lambda)|^2}{2\pi |1 + a_1 \exp(-i\lambda) + \cdots + a_p \exp(-ip\lambda)|^2}$$

Throughout the paper, σ is taken to be one. Four testing examples are:

Example 1. The AR(3) model with $a_1 = -1.5$, $a_2 = 0.7$ and $a_3 = -0.1$.

Example 2. The MA(4) model with $b_1 = -0.3$, $b_2 = -0.6$, $b_3 = -0.3$ and $b_4 = 0.6$.

Example 3: We generate the stationary time series X_t from $X_t = X_{t,1} + 4X_{t,2}$, where $X_{t,1}$ and $X_{t,2}$ are two independent time series, generated respectively from Examples 1 and 2.

Example 4: We use the AR12 model with $a_4 = -0.9$, $a_8 = -0.7$, $a_{12} = 0.63$ and the rest of the coefficients zero.

Put Figure 1 around here

Examples 1, 2 and 4 were used by Wahba (1980) to illustrate the smoothing spline method. We simulated 400 times with sample size $N = 250$ and present a typical estimated curve. More precisely, we present the curve \hat{m}_{LK} having the median performance (in terms of mean absolute deviation error) among 400 simulations. For comparison, we also present \hat{m}_{LS} using the same sample data. Figure 1 depicts the simulation results for estimation of the log-spectral density.

Put Figure 2 around here

We now illustrate the local likelihood method using sunspot data. The data consist of the annual average value of the daily index of the number of sunspots for the years 1749-1976. Figures 2 (a) and (b) give the estimated log-spectral and spectral densities. The automatic bandwidth selection rule gives bandwidths $\hat{h}_{\text{LS}, \text{OPT}} = 0.137$ and $\hat{h}_{\text{LK}, \text{OPT}} = (6/\pi^2)^{1/5} \hat{h}_{\text{LS}, \text{OPT}}$. Note that there are two peaks around frequencies 0 and $2\pi/10.8 \approx 0.58$. This indicates a 10.8-year cyclic pattern of the data.

3 Comparisons with other methods

Another classical method for estimating a spectral density is to smooth the periodogram directly, resulting in the spectral density estimator (instead of log-spectral density estimator) $\hat{f}_{\text{DLS}}(\lambda) = \sum_{k=1}^N w_k(\lambda) I^{(n)}(\lambda_k)$, where $w_k(\lambda)$ is the same local linear weight as in \hat{m}_{LS} . Since the spectral density is usually unsmooth, we apply the variable bandwidth selector of Fan and Gijbels (1995) to this estimator and obtain a fully automatic estimation method.

Put Table 1 around here

In this section, we compare the performance of three spectral density estimators \hat{m}_{LS} , \hat{m}_{LK} and \hat{f}_{DLS} based on 400 simulations. For a spectral density estimator \hat{f} , we define the Mean Absolute Deviation Error for the log-spectral density as

$$\text{MADE} = n_{\text{grid}}^{-1} \sum_{j=1}^{n_{\text{grid}}} |\log(\hat{f}(\lambda_j)) - \log(f(\lambda_j))|$$

and for the spectral density as

$$\text{SMADE} = n_{\text{grid}}^{-1} \sum_{j=1}^{n_{\text{grid}}} |\hat{f}(\lambda_j) - f(\lambda_j)|,$$

where $\lambda_j = j\pi/n_{\text{grid}}$ with $n_{\text{grid}} = 201$. For completeness, we also computed the corresponding Mean Average Square Errors for estimating the log-spectral density and the spectral density, denoting them respectively by MASE and SMASE. Figure 3 depicts the distributions of the mean absolute deviation errors based on 400 simulations. Table 1 summarizes the median for each of the distributions.

Put Figure 3 around here

From Figure 3 and Table 1, one can easily see that \hat{m}_{LK} performs better than \hat{m}_{LS} and $\log(\hat{f}_{\text{DLS}})$ when the optimal bandwidth, selected automatically by the procedures mentioned above for each estimator, is used. In particular, from Table 1, the median of the MADE and of the MASE for \hat{m}_{LK} are about 0.81 – 0.86 as large as those of \hat{m}_{LS} . This factor is comparable with the asymptotic result $(6/\pi^2)^{2/5} = 0.8195$. Even though $\exp(\hat{m}_{\text{LK}})$ performs better than \hat{f}_{DLS} in the above four examples, their asymptotic biases are not directly comparable while their asymptotic variance is the same.

The above simulations clearly indicate that the likelihood-based data-driven spectral density and log-spectral density estimators are the most appealing among their peers.

4 Sampling properties

In this section, we establish the asymptotic properties of the proposed local likelihood estimators. To simplify technical arguments, we consider the linear Gaussian process given by

$$X_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j}, \quad Z_j \sim \text{iid } N(0, \sigma^2).$$

The periodogram of this linear process is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} |A(\lambda)|^2, \quad \text{with } A(\lambda) = \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda}.$$

We make the following assumptions:

- (i) $\{X_t\}$ is a stationary process with $\sum_j |a_j|^2 < \infty$.
- (ii) The spectral density function $f(\cdot)$ is positive on $[0, \pi]$.
- (iii) The kernel function K is a symmetric probability density function and has a compact support.
- (iv) $(\log n)^4 h \rightarrow 0$ in such a way that $nh \rightarrow \infty$.

It follows from the first condition that the spectral density function has a bound second derivative.

For the local likelihood estimator \hat{m}_{LK} , we have the following results.

Theorem 1 *Under Conditions (i) – (iv), we have for each $0 < \lambda < \pi$,*

$$\sqrt{nh} \{ \hat{m}_{\text{LK}}(\lambda) - m(\lambda) - h^2 m''(\lambda) \mu_2(K)/2 + o(h^2) \} \xrightarrow{\mathcal{D}} N(0, \nu_0(K)\pi),$$

and for a boundary point $\lambda_n^* = ch$, we have

$$\sqrt{nh} \{ \hat{m}_{\text{LK}}(\lambda_n^*) - m(\lambda_n^*) - h^2 m''(0+) \mu_2(K, c)/2 + o(h^2) \} \xrightarrow{\mathcal{D}} N(0, \nu_0(K, c)\pi),$$

where with $s_{j,c} = \int_{-\infty}^c t^j K(t) dt$,

$$\mu_2(K, c) = \frac{s_{2,c}^2 - s_{1,c} s_{3,c}}{s_{0,c} s_{2,c} - s_{1,c}^2}, \quad \nu_0(K, c) = \frac{\int_{-\infty}^c (s_{2,c} - s_{c,1} t)^2 K^2(t) dt}{(s_{0,c} s_{2,c} - s_{1,c}^2)^2}.$$

A natural estimator for spectral density is $\hat{f}_{\text{LK}}(\lambda) = \exp\{\hat{m}_{\text{LK}}(\lambda)\}$. The following corollary describes its asymptotic behavior.

Corollary 1 Under Conditions (i) – (iv), we have for each $0 < \lambda < \pi$,

$$\sqrt{nh}\{\hat{f}_{\text{LK}}(\lambda) - f(\lambda) - h^2 m''(\lambda)f(\lambda)\mu_2(K)/2 + o(h^2)\} \xrightarrow{\mathcal{D}} N\left(0, \nu_0(K)f^2(\lambda)\pi\right),$$

and for a boundary point $\lambda_n^* = ch$, we have

$$\sqrt{nh}\{\hat{f}_{\text{LK}}(\lambda_n^*) - f(\lambda_n^*) - h^2 m''(0+)f(0+)\mu_2(K, c)/2 + o(h^2)\} \xrightarrow{\mathcal{D}} N\left(0, \nu_0(K, c)f^2(0+)\pi\right).$$

Remark 1: The results of Theorem 1 can easily be extended to general local polynomial fitting. More precisely, letting $\hat{m}_\nu(\lambda)$ be the estimated ν^{th} derivative of m based on the local polynomial fit of order p using the Whittle likelihood, then for $0 < \lambda < \pi$,

$$\begin{aligned} & \sqrt{nh^{2\nu+1}}\{\hat{m}_\nu(\lambda) - m(\lambda) - \int t^{p+1} K_{\nu,p}(t) dt \frac{m^{(p+1)}(\lambda)}{(p+1)!} h^{p+1-\nu} + o(h^{p+1-\nu})\} \\ & \xrightarrow{\mathcal{D}} N\left(0, \pi \int K_{\nu,p}^2(t) dt\right), \end{aligned}$$

where $K_{\nu,p}(t)$ is the equivalent kernel given in (5) of Fan, Heckman and Wand (1995). A similar result holds for the λ near the boundary.

Remark 2. For the smoothed periodogram \hat{f}_{DLS} , we have for each $0 < \lambda < \pi$,

$$\sqrt{nh}\{\hat{f}_{\text{DLS}}(\lambda) - f(\lambda) - h^2 f''(\lambda)\mu_2(K)/2 + o(h^2)\} \xrightarrow{\mathcal{D}} N\left(0, \nu_0(K)f^2(\lambda)\pi\right).$$

Note that $f''(\lambda) = f(\lambda)m''(\lambda) + f(\lambda)\{m'(\lambda)\}^2$. By Corollary 1, \hat{f}_{DLS} has the same asymptotic variance as that of \hat{f}_{LK} and has a larger bias than \hat{f}_{LK} when $m''(\lambda) > 0$. When $m''(\lambda) < 0$, it is hard to say which estimator has a larger bias.

Acknowledgments

Fan was partially supported by NSF Grants DMS-9203135 and DMS-9504414 and NSA Grant 96-1-0015. Kreutzberger was supported by a postdoctoral grant of the DFG (German Science Foundation). We thank Professor Young K. Truong and referees for their valuable comments that lead to significant improvement of presentation.

5 Proofs

By Theorem 10.3.1 of Brockwell and Davis (1991), we can decompose the periodogram as follows:

$$I^{(n)}(\lambda_k) = |A(\lambda_k)|^2 \cdot I_Z(\lambda_k) + R_{n,k} = f(\lambda_k)V_k + R_{n,k},$$

where $I_Z(\lambda) = \frac{1}{2\pi n} |\sum_{t=0}^{n-1} e^{-i\lambda t} Z_t|^2$ is the periodogram of the noise process $\{Z_t\}$ and $V_k = 2\pi\sigma^{-2}I_Z(\lambda_k)$. By the Gaussian assumption, V_1, \dots, V_N are i.i.d. exponentially distributed, where $N = [(n-1)/2]$.

Lemma 1 Under Condition (ii),

$$\max_{1 \leq k \leq N} |R_{n,k}| = O_p\left(\frac{\log n}{\sqrt{n}}\right).$$

The lemma follows directly from an explicit expression of $R_{n,k}$ given at page 347 of Brockwell and Davis (1991) and is proved by Kooperberg, Stone and Truong (1995b).

Proof of Theorem 2. The idea of this proof is to reduce the problem for the dependent data to that for the i.i.d. exponential distributions. We outline the key step of the proof.

Let $\hat{\boldsymbol{\beta}} = a_n^{-1}(\hat{a} - m(\lambda), h\{\hat{b} - m'(\lambda)\})^T$, where $a_n = (nh)^{-1/2}$ and (\hat{a}, \hat{b}) maximizes (2.5). Then, it can easily be seen that $\hat{\boldsymbol{\beta}}$ maximizes

$$\sum_{k=1}^N [-\exp\{Y_k - \bar{m}(\lambda, \lambda_k) - a_n \boldsymbol{\beta}^T \boldsymbol{\lambda}_k\} + Y_k - \bar{m}(\lambda, \lambda_k) - a_n \boldsymbol{\beta}^T \boldsymbol{\lambda}_k] K_h(\lambda_k - \lambda),$$

where

$$\bar{m}(\lambda, \lambda_k) = m(\lambda) + m'(\lambda)(\lambda_k - \lambda) \quad \text{and} \quad \boldsymbol{\lambda}_k = (1, (\lambda_k - \lambda)/h)^T.$$

Let

$$L_k(Y_k, \boldsymbol{\beta}) = -\exp\{Y_k - \bar{m}(\lambda, \lambda_k) - a_n \boldsymbol{\beta}^T \boldsymbol{\lambda}_k\} + Y_k - \bar{m}(\lambda, \lambda_k) - a_n \boldsymbol{\beta}^T \boldsymbol{\lambda}_k.$$

Then, $\hat{\boldsymbol{\beta}}$ maximizes

$$\ell_n(\boldsymbol{\beta}) = h \sum_{k=1}^N \{L_k(Y_k, \boldsymbol{\beta}) - L_k(Y_k, 0)\} K_h(\lambda_k - \lambda).$$

Let $Y'_k = m(\lambda_k) + \varepsilon_k$, the main term of (2.2). Then, we can write

$$\ell_n(\boldsymbol{\beta}) = \ell_{1,n}(\boldsymbol{\beta}) + U_n$$

where $\ell_{1,n}(\boldsymbol{\beta})$ is defined the same as $\ell_n(\boldsymbol{\beta})$ with Y_k replaced by Y'_k , and

$$U_n = -h \sum_{k=1}^N R_k \left[\exp\{-\bar{m}(\lambda, \lambda_k) - a_n \boldsymbol{\beta}^T \boldsymbol{\lambda}_k\} - \exp\{-\bar{m}(\lambda, \lambda_k)\} \right] K_h(\lambda_k - \lambda).$$

By using Taylor's expansion and Lemma 1, for each fixed $\boldsymbol{\beta}$,

$$U_n = O_p(h \cdot a_n \cdot n \cdot \log n / \sqrt{n}) = o_P(1).$$

Thus, we have

$$\ell_n(\boldsymbol{\beta}) = \ell_{1,n}(\boldsymbol{\beta}) + o_P(1). \tag{5.1}$$

Observe that $\ell_{1,n}(\boldsymbol{\beta})$ is the local likelihood based on $\{\exp(Y'_k), k = 1, \dots, N\}$ which are i.i.d exponentially distributed. It is a special case studied by Fan, Heckman and Wand (1995). As shown in Fan, Heckman and Wand (1995), for each $\boldsymbol{\beta}$ we have

$$\ell_{1,n}(\boldsymbol{\beta}) = \mathbf{W}_n^T \boldsymbol{\beta} + \frac{1}{2} \boldsymbol{\beta}^T A \boldsymbol{\beta} + o_P(1), \tag{5.2}$$

where $A = -\pi^{-1}\text{diag}\{1, \mu_2(K)\}$, and with $B = \pi^{-1}\text{diag}(\nu_0(K), \int t^2 K^2(t)dt)$

$$\mathbf{W}_n - a_n^{-1} \left\{ \frac{m''(x)}{2} h^2 \pi^{-1} (\mu_2(K), 0)^T + o(h^3) \right\} \xrightarrow{\mathcal{D}} N(0, B). \quad (5.3)$$

Combination of (5.1) and (5.2) leads to

$$\ell_n(\boldsymbol{\beta}) = \mathbf{W}_n^T \boldsymbol{\beta} + \frac{1}{2} \boldsymbol{\beta}^T A \boldsymbol{\beta} + o_P(1).$$

As noted before, $\mathcal{L}(a, b)$ is a concave function and so is $\ell_n(\boldsymbol{\beta})$. By the Quadratic Approximation Lemma given in Fan, Heckman and Wand (1995), we conclude that

$$\hat{\boldsymbol{\beta}} = A^{-1} \mathbf{W}_n + o_P(1).$$

The result follows from (5.3) and the last equality by taking its first component. This concludes the proof of Theorem 2.

Proof of Corollary 1. By using the Taylor's expansion and Theorem 1, we have

$$\begin{aligned} \hat{f}_{\text{LK}}(\lambda) - f(\lambda) &= \exp\{\hat{m}_{\text{LK}}(\lambda)\} - \exp\{m(\lambda)\} \\ &= \exp\{m(\lambda)\} \{\hat{m}_{\text{LK}}(\lambda) - m(\lambda)\} + O_p\left(\{\hat{m}_{\text{LK}}(\lambda) - m(\lambda)\}^2\right) \\ &= \exp\{m(\lambda)\} \{\hat{m}_{\text{LK}}(\lambda) - m(\lambda)\} + O_P\{h^4 + (nh)^{-1}\}. \end{aligned}$$

The conclusion follows.

References

- Beltrão, K.I. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *Journal of Time Series Analysis*, **8**, 21-36.
- 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*, second edition. Springer-Verlag, New York.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829 – 836.
- Davis, H.T. and Jones, R.H. (1968). Estimation of the innovation variance of a stationary time series. *Journal of the American Statistical Association*, **63**, 141-149.
- Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87**, 998 – 1004.
- Fan, J., Heckman, N.E. and Wand, M.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, **90**, 141-150.

- Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of Royal Statistical Society B*, **57**, 349 – 350.
- Fan, J. and Chen, J. (1997). One-step local quasi-likelihood estimation. *Manuscript*.
- Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics*, **20**, 121-145.
- Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statistical Sciences*, **8**, 120 – 143.
- Hurvich, C.M. and Beltrão, K.I. (1990). Cross-validatory choice of a spectral estimate and its connections with AIC. *Journal of Time Series Analysis*, **11**, 121-137.
- Jones, M.C. (1997). A variation on local linear regression. *Statistica Sinica*, to appear.
- Kooperberg, C., Stone, C.J. and Truong, Y.K. (1995a). Logspline estimation of a possibly mixed spectral distribution. *Journal of Time Series Analysis*, **16**, 359-389.
- Kooperberg, C., Stone, C.J. and Truong, Y.K. (1995b). Rate of convergence for logspline spectral density estimation. *Journal of Time Series Analysis*, **16**, 389-401.
- Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82**, 231 – 238.
- Müller, H.G. and Prewitt, K. (1992). Weak convergence and adaptive peak estimation for spectral densities, **20**, 1329-1349.
- Newton, H.J. and Pagano, M. (1983). A method for determining periods in time series. *Journal of American Statistical Association*, **78**, 152-157.
- Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association*, **89**, 600 - 610.
- Politis, D.N. and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α -mixing random variables with application to the problem of spectral density estimation. *The Annals of Statistics*, **20**, 1985-2007.
- Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press, London.
- Ruppert, D. and Wand, M.P. (1994). Multivariate weighted least squares regression. *The Annals of Statistics*, **22**, 1346 – 1370.
- Swanepoel, J.W. and van Wyk, J.W.J. (1986). The bootstrap applied to spectral density function estimation. *Biometrika*, **73**, 135-142.
- Stone, C.J. (1977). Consistent Nonparametric Regression. *The Annals of Statistics*, **5**, 595 – 645.
- Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, **75**, 122-132.

Whittle, P. (1962). Gaussian estimation in stationary time Series. *Bulletin of the International Statistical Institute*, **39**, 105-129.

Table 1: The Median of MADE, SMADE, MASE, SMASE based on 400 simulations

	\hat{m}_{LS}	\hat{m}_{LK}	$\log(\hat{f}_{DLS})$	$\exp(\hat{m}_{LS})$	$\exp(\hat{m}_{LK})$	\hat{f}_{DLS}
	MADE			SMADE		
Example 1	.1556	.1263	.2103	.3518	.2811	.4237
Example 2	.2327	.1991	.2198	.0586	.0485	.0472
Example 3	.2221	.1870	.2150	1.376	1.104	1.202
Example 4	.3484	.2809	.4498	4.451	3.429	4.630
	MASE			SMASE		
Example 1	.1960	.1909	.2674	1.090	.8417	1.137
Example 2	.2957	.2566	.3047	.0908	.0758	.0707
Example 3	.2835	.2445	.2950	2.067	1.638	1.822
Example 4	.4493	.3648	.6485	23.32	17.69	18.13

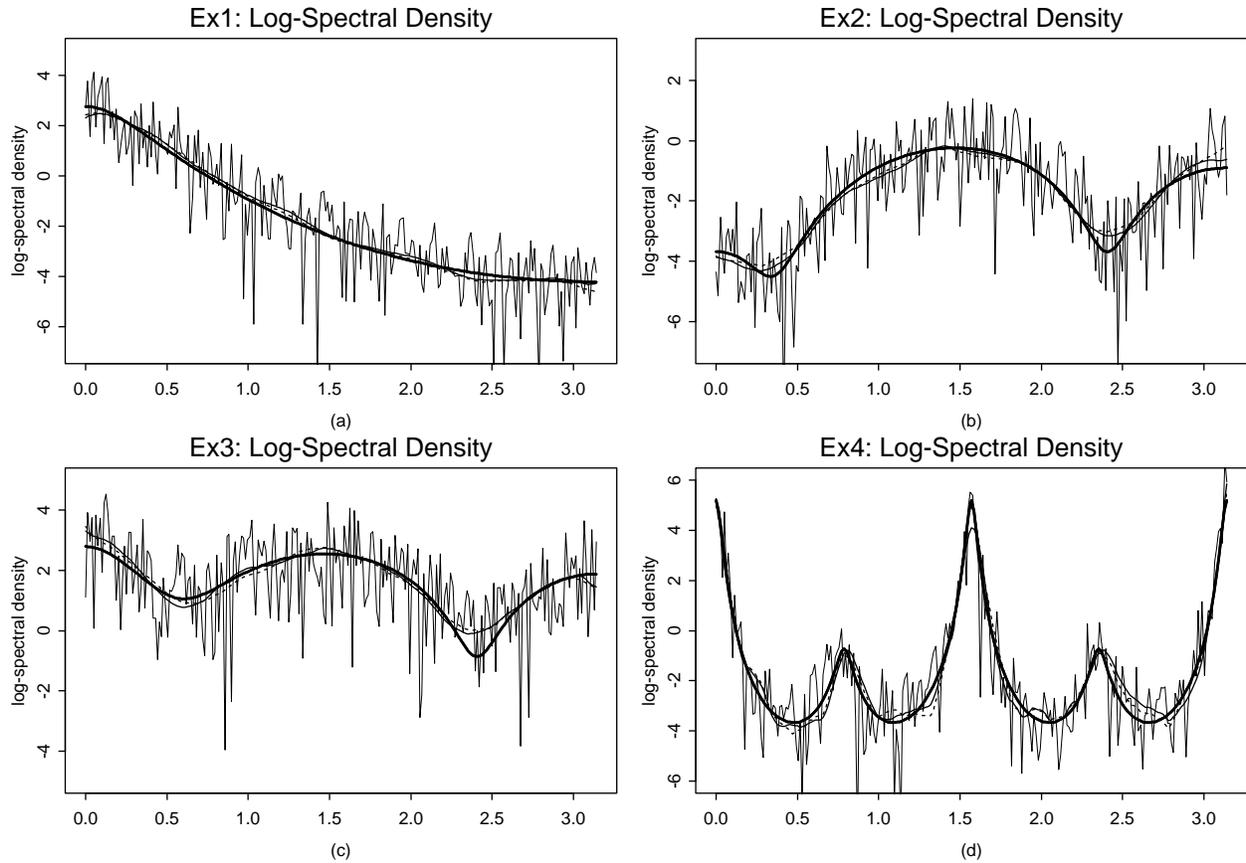


Figure 1: *Typical estimated log-spectral densities using the local likelihood method and the local least-squares method. Solid thick curves (from thickest to thinnest) are the true log-spectral densities, estimated log-spectral densities \hat{m}_{LK} , and log-periodogram plus 0.57221. Dashed curves are the estimated log-spectral densities $\hat{m}_{LS}(\cdot)$.*

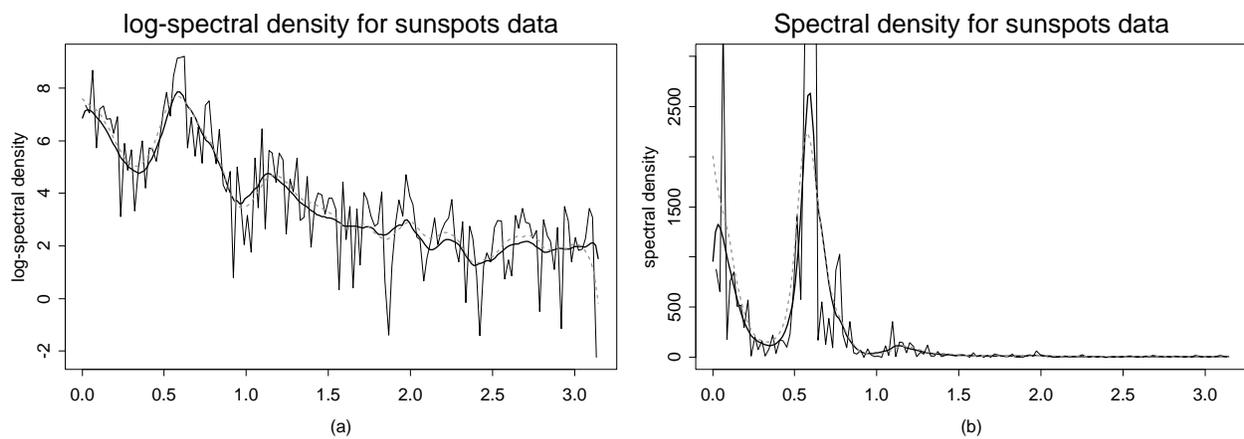


Figure 2: *Spectral density estimates for sunspot data. (a) estimated log-spectral density for the sunspots data; (b) estimated spectral density for the sunspots data. Thin solid curves — log-periodogram plus 0.57721 or periodogram; thick solid curves — local likelihood method; dashed curve — the smoothed periodogram using the least-squares method.*

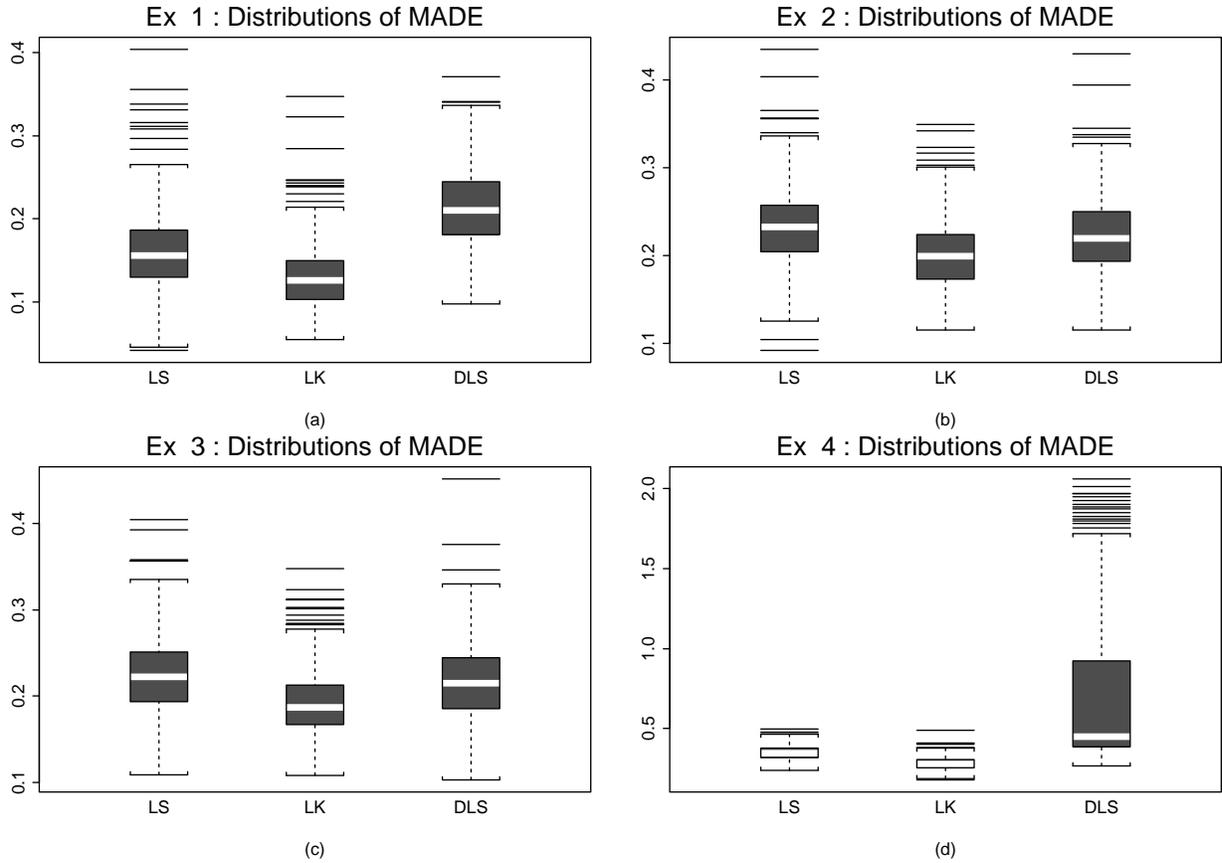


Figure 3: The boxplot of the distributions for the mean absolute deviation errors using models presented in Examples 1 – 4. From the left to right are the distributions for respectively \hat{m}_{LS} , \hat{m}_{LK} and $\log(\hat{f}_{DLS})$.