Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Signal Processing 84 (2004) 1255 – 1266

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Kernel smoothing of periodograms under Kullback–Leibler discrepancy

Jan Hannig, Thomas C.M. Lee*

Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA

Received 5 June 2003; received in revised form 31 March 2004

Abstract

Kernel smoothing on the periodogram is a popular nonparametric method for spectral density estimation. Most important in the implementation of this method is the choice of the bandwidth, or span, for smoothing. One idealized way of choosing the bandwidth is to choose it as the one that minimizes the Kullback–Leibler (KL) discrepancy between the smoothed estimate and the true spectrum. However, this method fails in practice, as the KL discrepancy is an unknown quantity. This paper introduces an estimator for this discrepancy, so that the bandwidth that minimizes the unknown discrepancy can be empirically approximated via the minimization of it. It is shown that this discrepancy estimator is consistent. Numerical results also suggest that this empirical choice of bandwidth often outperforms some other commonly used bandwidth choices. The same idea is also applied to choose the bandwidth for log-periodogram smoothing.

Keywords: Bandwidth selection; Kullback–Leibler discrepancy; Periodogram and log-periodogram smoothing; Relative entropy; Spectral density estimation

1. Introduction

Smoothing the periodogram or the log-periodogram is a popular method for performing nonparametric spectral density estimation. Various approaches have been proposed. These include spline smoothing (e.g., [9,13,19]), kernel smoothing (e.g., [4,10,11,18]) and wavelet techniques (e.g., [5,15,20]). The approach that this paper considers is kernel smoothing. Some appealing features of this approach are that it is simple to use, easy to understand and straightforward to interpret.

One important component of the kernel smoothing approach is the choice of the bandwidth for smoothing. This paper proposes a bandwidth selection method that attempts to locate the bandwidth that minimizes the unknown Kullback–Leibler (KL) discrepancy between the smoothed estimate and the true spectrum. The idea behind the proposed method is as follows. An estimator for the unknown KL discrepancy is first constructed for the class of spectrum estimates (i.e., kernel-smoothed periodograms) that this paper considers. It is shown that this discrepancy estimator is consistent under a commonly used model for periodograms. Then the bandwidth that minimizes this discrepancy estimator is chosen as the final bandwidth. As mentioned in [14], the rationale is that the bandwidth that minimizes the discrepancy estimator should also approximately minimize the unknown

* Corresponding author. Tel.: +1-970-491-5269; fax: +1-970-491-7895.

E-mail addresses: hannig@stat.colostate.edu (J. Hannig), tlee@stat.colostate.edu (T.C.M. Lee).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2004.04.007

discrepancy. Also, Chow [3, pp. 293–294] argues that the KL discrepancy is a better distance measure than the more commonly used squared distance measure. This point will be elaborated more later.

The rest of this paper is organized as follows. Section 2 provides some background material on the periodogram and the KL discrepancy. The proposed KL discrepancy estimator, together with some related previous work, are presented in Section 3. Section 4 briefly discusses the smoothing of log-periodogram. Some empirical and theoretical properties of the proposed method are reported in Sections 5 and 6, respectively. Conclusions are offered in Section 7, while technical details are deferred to the appendix.

2. Background

2.1. Periodogram smoothing

Suppose that $\{x_t\}$ is a real-valued, zero mean stationary process with unknown spectral density f, and that a finite-sized realization $x_0,\ldots,x_{2n-1}$ of $\{x_t\}$ is observed. The goal is to estimate f by using those observed x_ts. The periodogram is defined as

$$I(\omega)=\frac{1}{2\pi\times 2n}\left|\sum_{t=0}^{2n-1} x_t\exp(-\mathrm{i}\omega t)\right|^2,$$

$$\mathrm{i}=\sqrt{-1},\quad \omega\in[0,2\pi).$$

To simplify notation, write $\omega_j=2\pi j/(2n)$, $f_j=f(\omega_j)$ and $I_j=I(\omega_j)$. Since the spectral density f is symmetric about $\omega=\pi$, in the sequel the focus will be on f_j for $j=0,\ldots,n-1$. Also, as f is periodic with period 2π, one has $f_{-j}=f_j$ and $I_{-j}=I_j$ for $j=1,\ldots,n-1$.

A frequently adopted model for I_j is (e.g., see [4,10,15,17])

$$I_j=f_j\varepsilon_j,\quad j=0,\ldots,n-1,\tag{1}$$

where the ε_js are independent standard exponential random variables. Therefore $E(I_j)=f_j$ and $\mathrm{var}(I_j)=f_j^2$. Due to its unacceptably large variance, I_j is seldom used as an estimate of f_j.

One possible way for obtaining better estimates for f_j is to smooth the I_js. This paper considers the following kernel estimator for f_j:

$$\hat{f}_j=\sum_{m=-n}^{2n-1}K_h(\omega_m-\omega_j)I_m\Bigg/\sum_{l=-n}^{2n-1}K_h(\omega_l-\omega_j),$$
$$j=0,\ldots,n-1.\tag{2}$$

In the above $K_h(\cdot)=\frac{1}{h}K(\frac{\cdot}{h})$, where the kernel function K is (usually taken as) a symmetric probability density function and the bandwidth h is a nonnegative smoothing parameter that controls the amount of smoothing. Note that $\hat{f}_j$ is a function of h, but, for simplicity, this dependence is suppressed from its notation. It is well known that the choice of h is much more crucial than the choice of K (e.g., see [22]). Also, in most other kernel smoothing problems the limits of the two summations in (2) are 0 and $n-1$. However, since in the present setting boundary effects can be handled by periodic smoothing, the limits are changed from 0 and $n-1$ to $-n$ and $2n-1$, respectively.

The estimator $\hat{f}_j$ can also be interpreted as a weighted average of the I_js. It is because one could write

$$\hat{f}_j=\sum_{m=-n}^{2n-1}W_{m-j}I_m\quad\text{with}$$

$$W_{m-j}=\frac{K_h(\omega_m-\omega_j)}{\sum_{l=-n}^{2n-1}K_h(\omega_l-\omega_j)}.\tag{3}$$

Notice that the weights W_ms sum to unity.

2.2. Why KL discrepancy?

The KL discrepancy for measuring the distance between two probability density functions (pdfs) $g_1(x)$ and $g_2(x)$ is defined as

$$d(g_1,g_2)=\int g_1(x)\log\frac{g_1(x)}{g_2(x)}\,\mathrm{d}x$$

(e.g., see [2]). Note that $d(g_1,g_2)\neq d(g_2,g_1)$.

In order to use $d(g_1,g_2)$ for comparing a true spectrum f and an estimate $\hat{f}$, one needs to compare them frequency by frequency. At frequency ω_j, the pdf $g_f(x)$ corresponding to f is, under model (1), an exponential distribution with mean f_j. That is, $g_f(x)=\exp(-x/f_j)/f_j,\ x>0$. For $\hat{f}$, a natural candidate for the corresponding pdf is exponential with mean $\hat{f}_j$. Denote this pdf as $g_{\hat{f}}(x)$, and thus $g_{\hat{f}}(x)=\exp(-x/\hat{f}_j)/\hat{f}_j$, $x>0$. We choose to measure the distance between f and $\hat{f}$ at frequency ω_j with

$$d(g_{\hat{f}},g_f)=\int \frac{\exp(-x/\hat{f}_j)}{\hat{f}_j}\log\frac{\exp(-x/\hat{f}_j)/\hat{f}_j}{\exp(-x/f_j)/f_j}\,\mathrm{d}x$$

$$=\frac{\hat{f}_j}{f_j}-\log\frac{\hat{f}_j}{f_j}-1.$$

One could also use $d(g_f,g_{\hat{f}})$ instead of $d(g_{\hat{f}},g_f)$, but for the reason given in Appendix A, we choose $d(g_{\hat{f}},g_f)$. Therefore our overall KL discrepancy for measuring the distance between the whole spectrum f and $\hat{f}$ is

$$\Delta_{\mathrm{KL}}(\hat{f},f)=\frac{1}{n}\sum_{j=0}^{n-1}\left(\frac{\hat{f}_j}{f_j}-\log\frac{\hat{f}_j}{f_j}-1\right).$$

Our target is to choose the h that minimizes $\Delta_{\mathrm{KL}}(\hat{f},f)$. Notice that $\Delta_{\mathrm{KL}}(\hat{f},f)$ is an unknown quantity, so direct minimization is not possible.

A more frequently used discrepancy measure, based on the L_2 norm, is

$$\Delta_{L_2}(f,\hat{f})=\frac{1}{n}\sum_{j=0}^{n-1}(f_j-\hat{f}_j)^2.$$

However, as discussed in [3], $\Delta_{\mathrm{KL}}(\hat{f},f)$ is often a more relevant measure than $\Delta_{L_2}(f,\hat{f})$. It is because the former considers the distance for the whole distribution while the latter only considers the mean. Also, since $\Delta_{\mathrm{KL}}(\hat{f},f)$ has the same form as the theoretical deviance, it utilizes the asymptotic likelihood of the periodogram. Therefore, bandwidth selection methods that are targeting $\Delta_{\mathrm{KL}}(\hat{f},f)$ (or $E\{\Delta_{\mathrm{KL}}(\hat{f},f)\}$) seem to be more desirable than those that are targeting $\Delta_{L_2}(f,\hat{f})$ (or $E\{\Delta_{L_2}(f,\hat{f})\}$).

3. Estimating the KL discrepancy

This section presents the main contribution of this paper, namely, a consistent estimator for $\Delta_{\mathrm{KL}}(\hat{f},f)$. This estimator is denoted as $\hat{\Delta}_{h,k}$, and it is given by

$$\hat{\Delta}_{h,k}=\frac{1}{n}\sum_{j=0}^{n-1}\left\{\frac{2k}{\sum_{m=-k}^{k}I_{j+m}}\left(\hat{f}_j-\sum_{m=-k}^{k}W_{m-j}I_{j+m}\right)+\sum_{m=-k}^{k}W_{m-j}-\log\frac{\hat{f}_j}{I_j}+\gamma-1\right\},$$

where $\gamma\approx 0.577216$ is Euler's constant and k is a pre-specified positive integer parameter whose value is chosen independent of n. It will be shown in Theorem 1 that its effect is asymptotically negligible. We propose choosing h as the minimizer of $\hat{\Delta}_{h,k}$. Empirical properties of our estimator are reported in Section 5, while its consistency is established in Section 6.

3.1. Construction of $\hat{\Delta}_{h,k}$

This subsection outlines the construction of $\hat{\Delta}_{h,k}$. The goal is to seek an unbiased estimator for $\Delta_{\mathrm{KL}}(\hat{f},f)$.

First realize that estimating $\Delta_{\mathrm{KL}}(\hat{f},f)$ is equivalent to estimating the two quantities, $\log(\hat{f}_j/f_j)$ and $\hat{f}_j/f_j$, for all j. We estimate the former by $\log(\hat{f}_j/I_j)-\gamma$, as $E\{\log(\hat{f}_j/I_j)-\gamma\}=E\{\log(\hat{f}_j/f_j)\}$. The latter, due to the presence of $1/f_j$, poses a bigger challenge. It is because under model (1) $E(1/I_j)=\infty$, and so $\hat{f}_j/I_j$ cannot be used as a building block for estimating $\hat{f}_j/f_j$. To overcome this difficulty, we make use of the fact that if f is locally smooth, then $f_{j-k}\dots f_{j+k}$ are (approximately) identical for small k. This implies that the periodogram ordinates $I_{j-k},\dots,I_{j+k}$ are approximately independent and identically distributed as exponential with mean f_j. Therefore we have $E\{1/(\sum_{m=-k}^{k}I_{j+m})\}\approx 1/(2kf_j)$ and hence $1/f_j$ can be estimated by $2k/\sum_{m=-k}^{k}I_{j+m}$. Thus, there are two major advantages of introducing k: (1) it overcomes the problem of infinite expectation $E(1/I_j)=\infty$, and (2) it can be treated as a device for controlling the bias and variance of our estimator for $1/f_j$.

To proceed we decompose $\hat{f}_j/f_j$ into two parts:

$$\frac{\hat{f}_j}{f_j}=\sum_{m=-k}^{k}W_m\frac{I_{j+m}}{f_j}+\frac{(\hat{f}_j-\sum_{m=-k}^{k}W_mI_{j+m})}{f_j}.$$

We estimate the first part by its expectation, i.e., $\sum_{m=-k}^{k}W_m$. Since that the numerator of the second part and $\sum_{m=-k}^{k} I_{j+m}$ are independent, and that $E(2k/\sum_{m=-k}^{k} I_{j+m}) \approx 1/f_j$, we estimate the second part by $2k(\hat{f}_j - \sum_{m=-k}^{k} W_m I_{j+m})/(\sum_{m=-k}^{k} I_{j+m})$. Finally, by replacing $\log(\hat{f}_j/f_j)$ and $\hat{f}_j/f_j$ in the expression of $\Delta_{\mathrm{KL}}(\hat{f}, f)$ with these estimates, we obtain $\hat{\Delta}_{h,k}$.

3.2. Related work

A cross-validatory bandwidth selection criterion called CVLL has been studied by [1,7,8]. In particular, it is shown in [8] that CVLL is asymptotically equivalent to $E\{\Delta_{\mathrm{KL}}(f, \hat{f})\}$ (not $E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}$). However, it seems that $\Delta_{\mathrm{KL}}(f, \hat{f})$ (or $\Delta_{\mathrm{KL}}(\hat{f}, f)$) is to be preferred to $E\{\Delta_{\mathrm{KL}}(f, \hat{f})\}$ (or $E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}$). It is because $\Delta_{\mathrm{KL}}(f, \hat{f})$ measures the distance between f and $\hat{f}$ for the data set at hand, rather than for the average over all possible data sets.

More recently, [16] propose a generalized cross-validation (GCV)-based bandwidth selection method that also targets $\Delta_{\mathrm{KL}}(\hat{f}, f)$. The idea is to estimate $\Delta_{\mathrm{KL}}(\hat{f}, f)$ by cross-validating a gamma deviance function. However, no theoretical results are presented.

4. Log-periodogram smoothing

The above idea of KL discrepancy estimation can also be applied to choose the bandwidth for log-periodogram smoothing. The first step is to transform the multiplicative model (1) into an additive model by taking a logarithmic transform:

$$y_j = \log I_j + \gamma = \log f_j + \xi_j, \qquad l = 0, \ldots, n-1,$$

ξ_j : independent zero mean random variables with probability density function $p\xi(x) = \exp\{x - \gamma - \exp(x - \gamma)\}$.

Let $g = \log f$ be the log-spectrum and $\hat{g}$ be its kernel smoothing estimator; i.e., $\hat{g}_j = \sum_{m=-n}^{2n-1} W_{m-j} y_m$. It is straightforward to derive the following KL discrepancy for $\hat{g}$ and g:

$$\Delta'_{\mathrm{KL}}(\hat{g}, g) = \frac{1}{n}\sum_{j=0}^{n-1} \{g_j - \hat{g}_j + \exp(\hat{g}_j - g_j) - 1\}.$$

Using the same technique as in Section 3.1, we derived the following estimator $\hat{\Delta}'_{h,k}$ for $\Delta'_{\mathrm{KL}}(\hat{g}_h, g)$:

$$\hat{\Delta}'_{h,k} = \frac{1}{n}\sum_{j=0}^{n-1} \left\{ y_j - \hat{g}_j + C' \times \exp\left(\hat{g}_j - \sum_{m=-k}^{k} c_m y_{j+m} \right) - 1 \right\},$$

where

$$c_j = \frac{1}{2k+1} + W_j - \sum_{m=-k}^{k} \frac{W_m}{2k+1} \quad \text{and}$$

$$C' = \varepsilon^{\gamma} \prod_{m=-k}^{k} \frac{\Gamma(1 + W_m)}{\Gamma(1 + W_m - c_m)}.$$

It can be shown that, along the line of Theorems 1 and 2 below, $\hat{\Delta}'_{h,k}$ is consistent for $\Delta'_{\mathrm{KL}}(\hat{g}, g)$.

5. Finite sample properties

A numerical experiment was conducted for comparing the practical performance of the proposed bandwidth selection method with four other methods found in the literature. These other methods are (i) the CVLL and (ii) the GCV methods discussed in Section 3.2; (iii) the SES criterion which appeared in an unpublished report of Palmer (see [7]); and (iv) the $\hat{R}(p)$ criterion proposed by Lee [10]. The SES criterion uses cross-validation to estimate $E\{\Delta_{L_2}(f, \hat{f})\}$, while $\hat{R}(p)$ is an unbiased estimator for $E\{\Delta_{L_2}(f, \hat{f})\}$. Although these five methods are not targeting at the same discrepancy measure, a direct comparison of them would still be interesting. Throughout the whole experiment, k is set to 5 for the proposed method.

5.1. Setup

Four test examples and three different sample sizes were used. The three sample sizes were $n = 200$, 400 and 800, and the four test examples were from the ARMA(α, β) model

$$x_t + a_1 x_{t-1} + \cdots + a_\alpha x_{t-\alpha} = \tau_t + b_1 \tau_{t-1} + \cdots + b_\beta \tau_{t-\beta}, \qquad \tau_t \sim N(0,1)$$

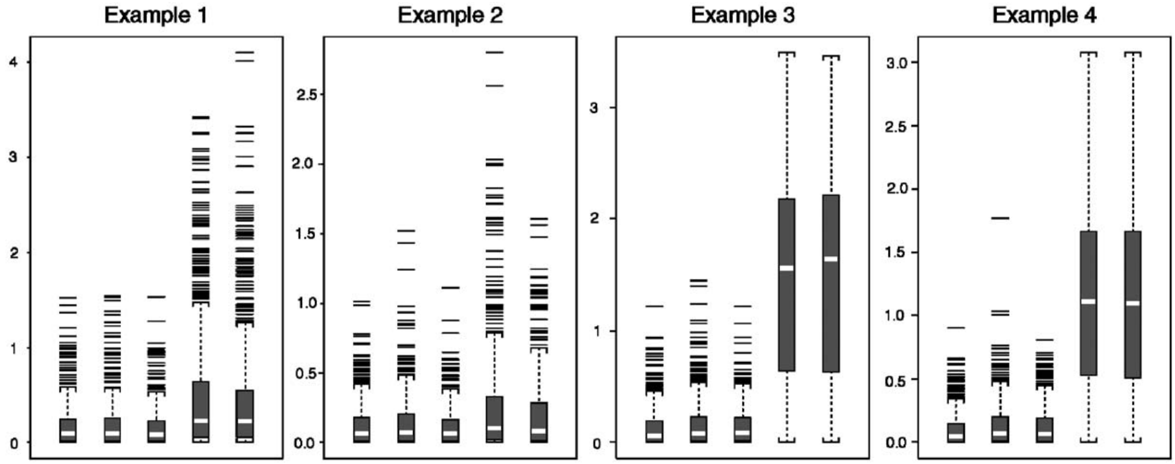

Fig. 1. Boxplots for various $\log R_{\mathrm{KL}}$ values. In each panel, the boxplots are corresponding, respectively, to, from left to right, the proposed method, the CVLL, GCV, SES and $\hat{R}(p)$ criteria.

with parameters given by: Example 1 was AR(3) with $a_1 = -1.5$, $a_2 = 0.7$, and $a_3 = -0.1$; Example 2 was AR(3) with $a_1 = 0.9$, $a_2 = 0.8$ and $a_3 = 0.6$; Example 3 was MA(3) with $b_1 = 0.9$, $b_2 = 0.8$ and $b_3 = 0.6$; and Example 4 was MA(4) with $b_1 = -0.3$, $b_2 = -0.6$, $b_3 = -0.3$ and $b_4 = 0.6$. These testing examples have previously been used by many authors; e.g., see [4,7,10,17,19]. The kernel function used was $K(x) = \frac{3}{4}(1 - x^2)$, $x \in [0, 1]$. It is the optimal kernel of order (0,2) derived in [6].

For each of the 12 combinations of test example and sample size, 500 independent series were simulated, and the corresponding periodograms were computed. Then, for each of these generated periodograms, the above five methods were applied to compute the corresponding bandwidths. In addition, two practically unobtainable bandwidths were also computed. They are h_{KL}, the bandwidth that minimizes $\Delta_{\mathrm{KL}}(\hat{f}, f)$, and h_{L_2}, the bandwidth that minimizes $\Delta_{L_2}(f, \hat{f})$.

Finally, the following two ratios were computed for every bandwidth h that was selected by any one of the five methods:

$$R_{\mathrm{KL}} = \frac{\Delta_{\mathrm{KL}}(\hat{f}, f)}{\Delta_{\mathrm{KL}}(\hat{f}_{h_{\mathrm{KL}}}, f)} \quad \text{and} \quad R_{L_2} = \frac{\Delta_{L_2}(f, \hat{f})}{\Delta_{L_2}(f, \hat{f}_{h_{L_2}})}$$

The first ratio R_{KL} is used to evaluate the performance of h with respect to $\Delta_{\mathrm{KL}}(\hat{f}, f)$: the smaller is its value, the better is the performance. The second ratio R_{L_2} has a similar interpretation, but is for $\Delta_{L_2}(f, \hat{f})$. Since CVLL is targeting at $E\{\Delta_{\mathrm{KL}}(f, \hat{f})\}$ but not $E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}$, we have also computed a similar third ratio that is based on $\Delta_{\mathrm{KL}}(f, \hat{f})$. However, as this third ratio gives the same empirical conclusions as for R_{KL}, the corresponding results will not be reported here.

For those cases that are associated with $n = 400$, boxplots of the log of the R_{KL} and R_{L_2} values are given in Figs. 1 and 2. Boxplots for other cases are somewhat similar and are hence omitted.

We also performed paired Wilcoxon tests to test the significance of the difference between the median R_{KL} (and R_{L_2}) values of any two methods. The significance level used was 1%, and the relative rankings, with 1 being the best, are listed in Tables 1 and 2. While ranking the methods in this manner is not perfectly legitimate, it does provide an indicator of the relative merits of the methods, and has also been used in other studies; e.g., see [12,13,21].

5.2. Empirical conclusions

With R_{KL} the averaged Wilcoxon test rankings for the proposed method, the CVLL, GCV, SES and $\hat{R}(p)$ criteria are 1.50, 2.54, 1.96, 4.75 and 4.25, respectively, while with R_{L_2} the averaged rankings

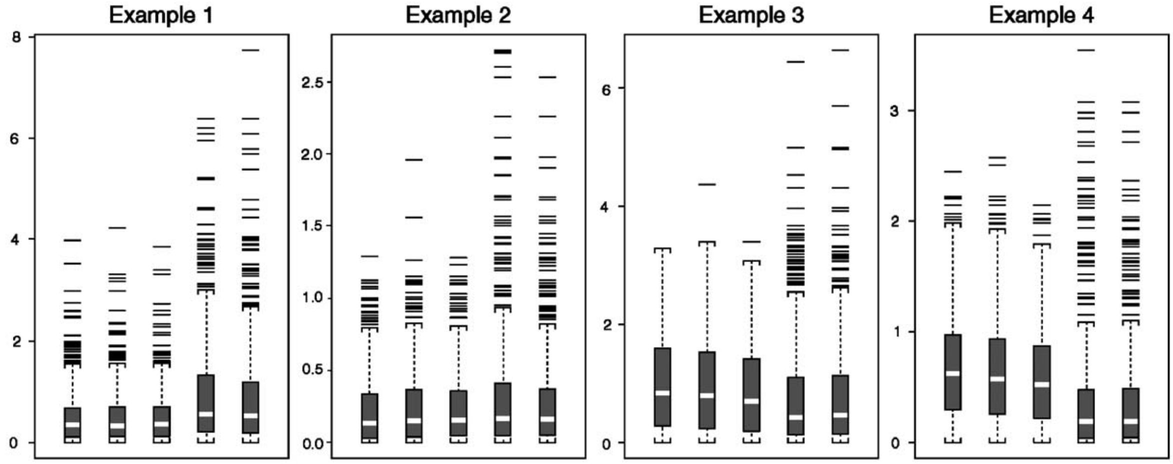

Fig. 2. Similar to Fig. 1 but for $\log R_{L_2}$.

Table 1
Wilcoxon rankings for comparing the R_{KL} values

Example	$n=200$					$n=400$					$n=800$				
	proposed	CVLL	GCV	SES	$\hat{R}(p)$	proposed	CVLL	GCV	SES	$\hat{R}(p)$	proposed	CVLL	GCV	SES	$\hat{R}(p)$
1	1.5	3	1.5	5	4	2.5	2.5	1	5	4	2.5	2.5	1	5	4
2	1	3	2	5	4	1.5	3	1.5	5	4	2	2	2	5	4
3	1	2.5	2.5	4.5	4.5	1	2.5	2.5	4.5	4.5	2	2	2	4.5	4.5
4	1	2.5	2.5	4.5	4.5	1	2.5	2.5	4.5	4.5	1	2.5	2.5	4.5	4.5

Table 2
Wilcoxon rankings for comparing the R_{L_2} values

Example	$n=200$					$n=400$					$n=800$				
	proposed	CVLL	GCV	SES	$\hat{R}(p)$	proposed	CVLL	GCV	SES	$\hat{R}(p)$	proposed	CVLL	GCV	SES	$\hat{R}(p)$
1	1.5	3	1.5	5	4	2	2	2	5	4	2	2	2	5	4
2	1	2.5	2.5	5	4	1	3.5	3.5	3.5	3.5	1.5	1.5	4	4	4
3	4.5	4.5	2.5	2.5	1	4.5	4.5	3	1.5	1.5	4	5	3	1.5	1.5
4	3.5	5	3.5	2	1	5	4	3	1.5	1.5	5	4	3	1.5	1.5

are 2.96, 3.46, 2.79, 3.17 and 2.63, respectively. Therefore, there are some evidence that the proposed method outperformed other methods if the target discrepancy is $\Delta_{\mathrm{KL}}(\hat{f}, f)$. It is interesting to see that even when the target discrepancy is $\Delta_{L_2}(f, \hat{f})$, the proposed method gave or shared the best results for Examples 1 and 2. Also, the widths of the boxplots in Figs. 1 and 2 seem to suggest that the three KL-based methods gave more stable performances than the two L_2-based methods. It is perhaps partially due to the fact that the KL discrepancy is more stable than the L_2 measure under heteroscedasticity.

5.3. Log-periodogram smoothing

We have also empirically investigated the use of the minimizer of $\hat{\Delta}'_{h,k}$ as the bandwidth for log-periodogram smoothing. We have compared this choice of bandwidth with other bandwidth selection methods that are targeting the L_2 measure $E\sum(g_j - \hat{g}_j)^2/n$. These other methods include cross-validation and the unbiased risk estimation approach of Lee [10]. Unlike the case for periodogram smoothing, the use of $\hat{\Delta}'_{h,k}$ only demonstrates a minor improvement over the other L_2-based methods. It is probably due to the fact that $E\Delta'_{\mathrm{KL}}(\hat{g}, g)$ and $E\sum(g_j - \hat{g}_j)^2/n$ are asymptotically equivalent.

6. Theoretical properties

We studied the theoretical properties of our estimator, under model (1), in terms of $\Theta = \hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)$. We summarize our results in two theorems. The first theorem provides nonasymptotic bounds for $E(\Theta)$ and $\mathrm{var}(\Theta)$, which can be used to show that, under some regularity conditions, both $E(\Theta)$ and $\mathrm{var}(\Theta)$ go to zero as $n \to \infty$. However, such a result is not very useful unless $\Delta_{\mathrm{KL}}(\hat{f}, f)$ goes to zero at a rate slower than Θ does. Our second theorem establishes the conditions under which $\Delta_{\mathrm{KL}}(\hat{f}, f)$ decays slower than Θ. That is, it enables us to show that $\hat{\Delta}_{h,k}$ is consistent for $\Delta_{\mathrm{KL}}(\hat{f}, f)$. The proofs of the theorems are given in Appendices B and C. Although these results are established under the approximation model (1), we believe that they do provide valuable insights to the performance of the proposed estimator.

Theorem 1. *Suppose that f is bounded away from 0 and ∞, the* kernel *K is unimodal, symmetrical and square integrable, and $k \geq 1$. Then, under model* (1),

$$|E(\Theta)| \leq \max_j (f_j) w\left(\frac{1}{f}, \frac{2\pi k}{n}\right) \tag{4}$$

and

$$\mathrm{var}(\Theta) \leq \frac{C_1}{n} + \frac{C_2 k}{n} w\left(\frac{1}{f}, \frac{2\pi k}{n}\right) + \frac{C_3 k^3}{n^3 h^2} + \frac{C_4}{n^2 h}, \tag{5}$$

where

$$C_1 \leq 30\,\frac{\max_j (f_j)^2}{\min_j (f_j)^2} + \pi^2/2, \qquad C_2 \leq 49\,\frac{\max_j (f_j)^2}{\min_j (f_j)^2},$$

C_3, C_4 are constants depending only on the kernel K, and w is the modulus of continuity (10).

One should notice that, from (4) and (5), the effect of the pre-chosen k on $E(\Theta)$ and $\mathrm{var}(\Theta)$ vanishes as n goes to ∞.

Theorem 2. *In addition to the assumptions stated in Theorem* 1, *we assume that f is twice differentiable, the derivative f' is bounded, $f''(x)$ does not vanish on some interval, and there is a nonincreasing sequence of h_n such that $h_n \to 0$ and $h_n = Cn^{-\alpha}$ for some constant C and $\alpha \in \{(0, \frac{1}{8}) \cup (\frac{1}{2}, \infty)\}$. Then*

$$\frac{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)}{\Delta_{\mathrm{KL}}(\hat{f}, f)} \to 0 \quad \textit{in probability.} \tag{6}$$

7. Conclusions

In this paper a bandwidth selection method for periodogram smoothing is proposed. The proposed method aims to choose the bandwidth that minimizes the KL discrepancy between the estimated and the true spectra. Some theoretical results are presented for supporting the use of this method. Results from numerical experiments seem to suggest that the proposed method is superior to four other existing bandwidth selection methods when the KL discrepancy is under consideration.

Appendix A. Why $d(g_{\hat{f}}, g_f)$ but not $d(g_f, g_{\hat{f}})$?

Using $d(g_{\hat{f}}, g_f)$ to measure the distance between f_i and $\hat{f}_j$ gives $\Delta_{\mathrm{KL}}(\hat{f}, f)$ as the overall KL discrepancy between f and $\hat{f}$, while using $d(g_f, g_{\hat{f}})$ gives

$$\Delta_{\mathrm{KL}}(f, \hat{f}) = \frac{1}{n}\sum_{j=0}^{n-1}\left(\frac{f_j}{\hat{f}_j} - \log\frac{f_j}{\hat{f}_j} - 1\right).$$

Now using the Taylor series approximation $y - \log y - 1 \approx (y-1)^2/2$ for y close to 1, we have

$$\Delta_{\mathrm{KL}}(\hat{f}, f) \approx \frac{1}{2n}\sum_{j=0}^{n-1}\left(\frac{f_j - \hat{f}_j}{f_j}\right)^2 \quad \text{and}$$

$$\Delta_{\mathrm{KL}}(f, \hat{f}) \approx \frac{1}{2n}\sum_{j=0}^{n-1}\left(\frac{f_j - \hat{f}_j}{\hat{f}_j}\right)^2.$$

Our belief is that $\Delta_{\mathrm{KL}}(\hat{f}, f)$ is a better measure to use. It is because in the above approximation it uses a fixed quantity, the denominator term f_j, to adjust for the variance $f_j - \hat{f}_j$ for different ω_j, while $\Delta_{\mathrm{KL}}(f, \hat{f})$ uses a random quantity $\hat{f}_j$. Also, our numerical work shows that the behaviors of $\Delta_{\mathrm{KL}}(\hat{f}, f)$ and $\Delta_{\mathrm{KL}}(f, \hat{f})$ are remarkably similar for h that are not far from the optimal value.

Appendix B. Proof of Theorem 1

To simplify notation, denote $\hat{f}_j^k = \sum_{m=-k}^{k} W_m I_{j+m}$. Also, decompose Θ as $\Theta = \frac{1}{n}\sum_{j=0}^{n-1} \Theta_j$, where

$$\Theta_j = \log I_j - (\log f_j - \gamma) + \frac{2k(\hat{f}_j - \hat{f}_j^k)}{\sum_{m=-k}^{k} I_{j+m}} + \sum_{m=-k}^{k} W_m - \frac{\hat{f}_j}{f_j}.$$

Proof of (4). By taking the expectation of Θ_j and using the fact that the observations are independent, we get

$$E(\Theta_j) = E(\hat{f}_j - \hat{f}_j^k)E\left(\frac{2k}{\sum_{m=-k}^{k} I_{j+m}}\right) + \sum_{m=-k}^{k} W_m - \frac{E(\hat{f}_j)}{f_j}. \tag{B.1}$$

Denote $x_j = \min_j(f_{j-k}, \ldots, f_{j+k})$ and $y_j = \max_j(f_{j-k}, \ldots, f_{j+k})$. Combining

$$\frac{1}{y_j} \leqslant E\left\{2k\left(\sum_{m=-k}^{k} I_{j+m}\right)^{-1}\right\} \leqslant \frac{1}{x_j} \tag{B.2}$$

and (B.1) we conclude

$$|E(\Theta_j)| \leqslant E(\hat{f}_j) w\left(\frac{1}{f}, \frac{2\pi k}{n}\right) \leqslant \max_j (f_j) w\left(\frac{1}{f}, \frac{2\pi k}{n}\right), \tag{B.3}$$

where w is the modulus of continuity defined for any continuous function g:

$$w(g, \varepsilon) = \sup\{|g(x) - g(y)|;\ |x - y| \leqslant \varepsilon\}. \tag{B.4}$$

Eq. (4) now follows from averaging (B.3) over j. □

Proof of (5). We first split the term $\Theta_j = X_j - Y_j + Z_j + \sum_{m=-k}^{k} W_m$, where

$$X_j = \log I_j - (\log f_j - \gamma), \qquad Y_j = \frac{\hat{f}_j}{f_j},$$

$$Z_j = \frac{2k(\hat{f}_j - \hat{f}_j^k)}{\sum_{m=-k}^{k} I_{j+m}}.$$

Firstly,

$$\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n}\Theta_j\right) \leqslant 3\,\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) + 3\,\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right) + 3\,\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} Z_j\right) \tag{B.5}$$

and under model (1)

$$\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{\pi^2}{6n}. \tag{B.6}$$

Secondly,

$$\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}\left(\frac{\hat{f}_i}{f_i}, \frac{\hat{f}_j}{f_j}\right)$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\operatorname{cov}\left(\sum_{p=-n}^{2n-1} W_{p-i}I_p,\sum_{q=-n}^{2n-1} W_{q-j}I_q\right)}{f_i f_j}$$

$$\leqslant \frac{1}{n^2 \min_j(f_j)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=-n}^{2n-1} W_{p-i}W_{p-j}\operatorname{var}(I_p)$$

$$\leqslant \frac{3\max_j(f_j)^2}{n\min_j(f_j)^2}. \quad (B.7)$$

Thirdly, fix $i \leqslant j$. In order to be able to estimate the last term we need to calculate $\operatorname{cov}(Z_i, Z_j)$.

Define the following index sets:

$$I_{i,j,k} = \{x \in \mathbb{Z}:\ -k \leqslant x \leqslant \min(k, j-i-k-1)\},$$

$$J_{i,j,k} = \{x \in \mathbb{Z}:\ -\min(k, j-i-k-1) \leqslant x \leqslant k\},$$

$$K_{i,j,k} = \{x \in \mathbb{Z}:\ -n \leqslant x \leqslant 2n-1,\ |x-i| > k,\ |x-j| > k\}$$

and calculate

$$\operatorname{cov}(Z_i, Z_j) =\operatorname{cov}\left(\frac{2k\sum_{m\in J_{i,j,k}} W_{j-i+m}I_{j+m}}{\sum_{m=-k}^{k} I_{i+m}}, \frac{2k\sum_{m\in I_{i,j,k}} W_{i-j+m}I_{i+m}}{\sum_{m=-k}^{k} I_{j+m}}\right) + \sum_{p\in K_{i,j,k}}\sum_{q\in K_{i,j,k}}\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}\right). \quad (B.8)$$

Notice,

$$\frac{2k\sum_{m\in J_{i,j,k}} W_{j-i+m}I_{j+m}}{\sum_{m=-k}^{k} I_{j+m}} \leqslant 2k \max_{m\in J_{i,j,k}} W_{j-i+m} = 2kW_{\max(k+1, j-i-k)},$$

where the last equality follows from the fact that the smoothing kernel K is symmetric unimodal.

Thus,

$$\operatorname{cov}\left(\frac{2k\sum_{m\in J_{i,j,k}} W_{j-i+m}I_{j+m}}{\sum_{m=-k}^{k} I_{i+m}}, \frac{2k\sum_{m\in I_{i,j,k}} W_{i-j+m}I_{i+m}}{\sum_{m=-k}^{k} I_{j+m}}\right) \leqslant 4k^2 W^2_{\max(k+1,|j-i|-k)}.$$

This inequality remains valid also for $i > j$ by symmetry. Further notice that

$$W_j \approx \frac{\pi K(\pi j/nh)}{nh} \leqslant \frac{\pi K(0)}{nh}, \sum_{m=-\infty}^{\infty} W_m^2 \approx \frac{\pi\int K^2(\omega)\,\mathrm{d}\omega}{nh}. \quad (B.9)$$

Therefore there are constants C_3, C_4 depending only on the kernel K, such that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov} \times \left(\frac{2k\sum_{m\in J_{i,j,k}} W_{j-i+m}I_{j+m}}{\sum_{m=-k}^{k} I_{i+m}}, \frac{2k\sum_{m\in I_{i,j,k}} W_{i-j+m}I_{i+m}}{\sum_{m=-k}^{k} I_{j+m}}\right) \leqslant \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 4k^2 W^2_{\max(k+1,|j-i|-k)} \leqslant \frac{(2k)^3}{n} W^2_{k+1} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{m=-\infty}^{\infty} W_m^2 \leqslant \frac{C_3 k^3}{n^3h^2} + \frac{C_4}{n^2 h}. \quad (B.10)$$

To estimate the second term on the right-hand side of (B.8), first recall the Eq. (B.2), conclude

$$\operatorname{var}\left(\frac{2k}{\sum_{m=-k}^{k} I_{j+k}}\right) \leqslant \frac{1}{x_j^2(2k+1)} + \frac{1}{x_j^2} - \frac{1}{y_j^2} \leqslant \frac{1}{\min_j(f_j)^2}\left\{\frac{1}{2k+1} + 2w\left(\frac{1}{f}, \frac{2\pi k}{n}\right)\right\},$$

and realize that $|\operatorname{cov}(X, Y)| \leqslant \sqrt{\operatorname{var}(X)\operatorname{var}(Y)}$ for any random variables X, Y. Then fix $p, q \in K_{i,j,k}$ and calculate

$$\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}\right)$$

$$=E\left\{\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p, I_q\right)\right\}$$

$$+\operatorname{cov}\left\{E\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}|I_p, I_q\right),\right.$$

$$\left. E\left(\frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p, I_q\right)\right\}. \qquad \text{(B.11)}$$

If $|i-j|>2k$ we have

$$\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p, I_q\right)=0.$$

Similarly, if $|i-j|\leqslant 2k$

$$E\left\{\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p, I_q\right)\right\}$$

$$\leqslant \begin{cases} W_{p-i}W_{q-j}f_pf_q\,\dfrac{1}{\min_j(f_j)^2} \\ \quad\times\left\{\dfrac{1}{2k+1}+2w\left(\dfrac{1}{f}, \dfrac{2\pi k}{n}\right)\right\} & \text{if } p\neq q, \\ W_{p-i}W_{q-j}2f_p^2\,\dfrac{1}{\min_j(f_j)^2} \\ \quad\times\left\{\dfrac{1}{2k+1}+2w\left(\dfrac{1}{f}, \dfrac{2\pi k}{n}\right)\right\} & \text{if } p=q \end{cases}$$

and

$$\sum_{p\in K_{i,j,k}}\sum_{q\in K_{i,j,k}}\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}\right)$$

$$\leqslant \frac{2\max_j(f_j)^2}{\min_j(f_j)^2}\left\{\frac{1}{2k+1}+2w\left(\frac{1}{f}, \frac{2\pi k}{n}\right)\right\}.$$

Since there is no more than $(4k+1)n$ pairs of (i,j) satisfying $|i-j|\leqslant 2k$, we conclude

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p\in K_{i,j,k}}\sum_{q\in K_{i,j,k}}\operatorname{cov}\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}, \frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}\right)$$

$$\leqslant \frac{2\max_j(f_j)^2}{\min_j(f_j)^2}\left\{\frac{2}{n}+\frac{8k+2}{n}\,w\left(\frac{1}{f}, \frac{2\pi k}{n}\right)\right\}. \qquad \text{(B.12)}$$

Finally, estimate the second part of (B.11). If $p\neq q$,

$$\operatorname{cov}\left\{E\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}|I_p, I_q\right),\right.$$

$$\left. E\left(\frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p, I_q\right)\right\}=0.$$

Similarly if $p=q$,

$$\operatorname{cov}\left\{E\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}|I_p\right), E\left(\frac{2kW_{q-j}I_q}{\sum_{m=-k}^{k} I_{j+k}}|I_p\right)\right\}$$

$$\leqslant \frac{1}{x_ix_j}\,W_{p-i}W_{p-j}f_p^2.$$

From here we conclude

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p\in K_{i,j,k}}\sum_{q\in K_{i,j,k}}\operatorname{cov}\left\{E\left(\frac{2kW_{p-i}I_p}{\sum_{m=-k}^{k} I_{i+k}}|I_p\right),\right.$$

$$\left. E\left(\frac{2kW_{p-j}I_p}{\sum_{m=-k}^{k} I_{j+k}}|I_p\right)\right\}$$

$$\leqslant \frac{1}{n^2}\sum_{p\in K_{i,j,k}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\max_j(f_j)^2}{\min_j(f_j)^2}\,W_{p-i}W_{p-j}$$

$$\leqslant \frac{3\max_j(f_j)^2}{n\min_j(f_j)^2}. \qquad \text{(B.13)}$$

Combining (B.5) to (B.8), and (B.10) to (B.13) we obtain (5). □

Appendix C. Proof of Theorem 2

We will first show that under our assumptions

$$E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}\geqslant D_1h_n^4+\frac{D_2}{nh} \quad \text{eventually.} \qquad \text{(C.1)}$$

We need two inequalities to prove (C.1). Define $l(y)=(y-1)-\log(y)$ and hence $\Delta_{\mathrm{KL}}(\hat{f}, f)=$

$n^{-1}\sum_{j=0}^{n} l(\hat{f}_j/f_j)$. By applying the Taylor approximation $l(y) \approx \frac{1}{2}(y-1)^2$ to $l(\hat{f}_j/f_j)$ (as in Appendix A) and using the assumption that f is bounded away from 0 and ∞, we obtain our first inequality:

$$C\left(\frac{\hat{f}_j - f_j}{f_j}\right)^2 \leqslant l\left(\frac{\hat{f}_j}{f_j}\right) \quad \text{with}$$

$$C = \frac{1}{2}\frac{\min(f_j)}{\max(f_j)}. \tag{C.2}$$

We begin deriving the second inequality by calculating

$$E(\hat{f}_j - f_j)^2 = \mathrm{var}(\hat{f}_j) + \{E(\hat{f}_j) - f_j\}^2. \tag{C.3}$$

If $f''(\omega_j) \neq 0$ then there is $D(\omega_j)$ such that eventually

$$\{E(\hat{f}_j) - f_j\}^2 = \left\{\sum_{m=-n}^{2n-1} W_{m-j} f_m - f_j\right\}^2 \geqslant D(\omega_j) h_n^4 \tag{C.4}$$

and by (B.9)

$$\mathrm{var}(\hat{f}_j) = \sum_{m=-n}^{2n-1} W_{m-j}^2 \,\mathrm{var}(I_m) \geqslant \frac{\min_j (f_j)^2 C_4'}{nh}. \tag{C.5}$$

Thus by combining (C.2), (C.3), averaged (C.4) and (C.5) we get (C.1).

Similar calculations as in the proof of (5) show that

$$\mathrm{var}\{\Delta_{\mathrm{KL}}(\hat{f}, f)\} = \mathrm{O}\left(\frac{1}{n}\right). \tag{C.6}$$

Also recall that since f' is bounded the relations (4) and (5) imply

$$E\{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)\}^2 = \mathrm{O}\left(\frac{1}{n}\right) \tag{C.7}$$

Combining (C.1), (C.6) and (C.7) we see that under the conditions stated in Theorem 2 $n^{-1/2} = \mathrm{o}(h_n^4 + (nh)^{-1})$. Therefore, we can find r_n such that

$$E\{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)\}^2 = \mathrm{o}(r_n^2),$$

$$\mathrm{var}\{\Delta_{\mathrm{KL}}(\hat{f}, f)\} = \mathrm{o}(r_n^2) \quad \text{and}$$

$$r_n = \mathrm{o}[E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}]. \tag{C.8}$$

Fix $\varepsilon > 0$ and calculate

$$P\left\{\left|\frac{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)}{\Delta_{\mathrm{KL}}(\hat{f}, f)}\right| > \varepsilon\right\} < P\left\{\left|\frac{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)}{r_n}\right| > \varepsilon\right\} + P\{\Delta_{\mathrm{KL}}(\hat{f}, f) < r_n\}.$$

By (C.8), the Markov's and Chebyshev's inequalities we get

$$P\left\{\left|\frac{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)}{r_n}\right| > \varepsilon\right\} < \frac{E\{\hat{\Delta}_{h,k} - \Delta_{\mathrm{KL}}(\hat{f}, f)\}^2}{\varepsilon^2 r_n^2} \to 0,$$

$$P\{\Delta_{\mathrm{KL}}(\hat{f}, f) < r_n\} < P[|\Delta_{\mathrm{KL}}(\hat{f}, f) - E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}| > E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\} - r_n] < \frac{\mathrm{var}\{\Delta_{\mathrm{KL}}(\hat{f}, f)\}}{[E\{\Delta_{\mathrm{KL}}(\hat{f}, f)\} - r_n]^2} \to 0.$$

This finishes the proof of (6). □

References

[1] K. Beltrão, P. Bloomfield, Determining the bandwidth of a kernel spectrum estimate, J. Time Ser. Anal. 8 (1987) 21–39.

[2] K.P. Burnham, D.R. Anderson, Model Selection and Inference: A Practical Information–Theoretic Approach, Springer, New York, 1998.

[3] G.C. Chow, Econometrics, McGraw-Hill, New York, 1983.

[4] J. Fan, E. Kreutzberger, Automatic local smoothing for spectral density estimation, Scand. J. Statist. 25 (1998) 359–369.

[5] H.-Y. Gao, Choice of thresholds for wavelet shrinkage estimate of the spectrum, J. Time Ser. Anal. 18 (1997) 231–251.

[6] T. Gasser, H.-G. Müller, V. Mammitzsch, Kernels for nonparametric curve estimation, J. Roy. Statist. Soc. Ser. B 47 (1985) 238–252.

[7] C.M. Hurvich, Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods, J. Amer. Statist. Assoc. 80 (1985) 933–940.

[8] C.M. Hurvich, K. Beltrão, Cross-validatory choice of a spectrum estimate and its connections with AIC, J. Time Ser. Anal. 11 (1990) 121–137.

[9] C. Kooperberg, C.J. Stone, Y.K. Troung, Logspline estimation of a possibly mixed spectral distribution, J. Time Ser. Anal. 16 (1995) 359–388.

[10] T.C.M. Lee, A simple span selector for periodogram smoothing, Biometrika 84 (1997) 965–969.

[11] T.C.M. Lee, A stabilized bandwidth selection method for kernel smoothing of the periodogram, Signal Processing 81 (2001) 419–430.

[12] T.C.M. Lee, Tree-based wavelet regression for correlated data using the minimum description length principle, Austral. New Zealand, J. Statist. 44 (2002) 23–39.

[13] T.C.M. Lee, T.F. Wong, Nonparametric log-spectrum estimation using disconnected regression splines and genetic algorithms, Signal Processing 83 (2003) 79–90.

[14] H. Linhart, W. Zucchini, Model Selection, Wiley, New York, 1986.

[15] P. Moulin, Wavelet thresholding techniques for power spectrum estimation, IEEE Trans. Signal Process. 42 (1994) 3126–3136.

[16] H.C. Ombao, J.A. Raz, R.L. Strawderman, R. von Sachs, A simple generalised crossvalidation method of span selection for periodogram smoothing, Biometrika 88 (2001) 1186–1192.

[17] Y. Pawitan, F. O'Sullivan, Nonparametric spectral density estimation using penalized Whittle likelihood, J. Amer. Statist. Assoc. 89 (1994) 600–610.

[18] K.S. Riedel, A. Sidorenko, Adaptive smoothing of the log-spectrum with multiple tapering, IEEE Trans. Signal Process. 44 (1996) 1794–1800.

[19] G. Wahba, Automatic smoothing of the log periodogram, J. Amer. Statist. Assoc. 75 (1980) 122–132.

[20] A.T. Walden, D.B. Percival, E.J. McCoy, Spectrum estimation by wavelet thresholding of multitaper estimators, IEEE Trans. Signal Process. 46 (1998) 3153–3165.

[21] M.P. Wand, A comparison of regression spline smoothing procedures, Comput. Statist. 15 (2000) 443–462.

[22] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.