Journal of the Korean
Statistical Society
Vol. 23, No. 1, 1994

An Automatic Spectral Density Estimate†

Byeong U. Park[1], Sinsup Cho[1], Kee H. Kang[1]

ABSTRACT

This paper concerns the problem of estimating the spectral density function in the analysis of stationary time series data. A kernel type estimate is considered, which entails choice of bandwidth. A data-driven bandwidth choice is proposed, and it is obtained by plugging some suitable estimates into the unknown parts of a theoretically optimal choice. A theoretical justification is given for this choice in terms of how far it is from the theoretical optimum. Furthermore, an empirical investigation is done. It shows that the data-driven choice yields a reliable spectrum estimate.

KEYWORDS: Kernel estimate, Bandwidth, Spectral density

1. INTRODUCTION

Spectral density has an important role in the analysis of stationary time series. See Brillinger (1981) and Priestly (1981) for its interpretation and relation to the autocorrelation function. Although fitting an autoregression

[1] Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea.

† This research was supported in part by KOSEF 911-0105-014-2.

is most popular as a method of estimating the spectral density, it is based on fitting a particular model and so is defective when the fitted model is inappropriate. As an alternative, a kernel estimate may be used and it is obtained by smoothing the periodogram with a spectral window (kernel). But a difficulty arises in practical use of this estimate since it involves a parameter, called bandwidth, that determines the smoothness of the result, a parameter whose value must be chosen.

There are only a few literature on the bandwidth selection problem in kernel spectrum estimation. Those include, among others, Robinson (1991) for asymptotic theory of some kernel spectrum estimates in the presence of data-dependent bandwidths, Hurvich (1985) and Beltrão and Bloomfield (1987) for application of cross-validation methods. In this paper, a data-driven choice of the bandwidth is proposed. It is based on the asymptotic mean integrated squared error of the spectrum estimate.

In the next section, we describe the estimate and state two theorems. The first one gives a theoretically optimal bandwidth and hence motivates the data-driven bandwidth choice, the second shows how close these two bandwidths. Section 3 contains the results of a simulation study of the small sample properties of the estimate. It compares the two bandwidths in terms of estimating f, and reveals that there is no loss of efficiency (there is gain in some cases) in the use of the data-driven bandwidth. Technical proofs are deferred to Section 4.

2. THE ESTIMATE AND LARGE SAMPLE RESULTS

Let $X(t),\ \ t = 0, \pm 1, \ldots,$ be a stationary time series with zero mean and the spectral density

$$f(\lambda) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} exp(-i\lambda s) E\{X(0)X(s)\}, \quad -\infty < \lambda < \infty.$$

A kernel spectrum estimate is given by

$$\hat{f}_h(\lambda) = K_h * I_n(\lambda), \quad -\infty < \lambda < \infty. \tag{2.1}$$

Here and below $I_n(\lambda)$ denotes the periodogram of the data, defined by

$$I_n(\lambda) = (2\pi n)^{-1}\Big|\sum_{s=0}^{n-1} exp(-i\lambda s)X(s)\Big|^2, \; -\infty < \lambda < \infty,$$

$K_h(\cdot) = K(\cdot/h)/h$, K is called kernel, $h = h_n$ bandwidth, and $*$ denotes convolution. Throughout this paper, K is taken a nonnegative kernel with bounded support $[-1,1]$.

In pratice, a discretized version of (2.1) may be used for fast computation. Let $\delta = h/M$ and $B_j = [(j-1/2)\delta, \; (j+1/2)\delta]$ where M is a positive integer which determines the amount of discretization errors. Let $r(x)$ denote the 'rounded point of x/δ', defined by $r(x) = j$ if and only if $x \in B_j$. Then the estimate (2.1) can be approximated by

$$h^{-1}\int K(\frac{r(\lambda)-r(x)}{M})I_n(r(x)\delta)dx.$$

Since K has support $[-1,1]$, it is further approximated by

$$\tilde{f}_h(\lambda) = \frac{1}{M}\sum_{l=1-M}^{M-1} K(\frac{l}{M})I_n((r(\lambda)+l)\delta). \tag{2.2}$$

The essential idea of the discretization is that one replaces x by $\delta r(x)$.

Although (2.2) has some pratical advantages, we will consider (2.1) in the asymptotic analysis. For, it is simpler to analyze and the two estimates are arbitrarily close to each other as M goes to infinity. However, we will return to (2.2) in the simulation study below.

An appealing approach to the problem of selecting h is to consider an error criterion, such as the Mean Integrated Squared Error,

$$MISE(h) = E\int_{-\pi}^{\pi}\{\hat{f}_h(\lambda) - f(\lambda)\}^2 d\lambda.$$

To get an asymptotic representation of $MISE(h)$, we need the following assumptions on the smoothness of power spectra :

(A.1) $$\sum_{u_1,\ldots,u_{p-1}=-\infty}^{\infty} (1+|u_j|)|C_p(u_1,\ldots,u_{p-1})| < \infty,$$

denotes the p-th order cumulant of $X(0), X(u_1), \ldots, X(u_{p-1})$.

(A.2) The spectral density f has smoothness of order $\zeta > 2m$, i.e., there is a constant $c > 0$ such that, for all x and y

$$|f^{(l)}(x) - f^{(l)}(y)| \le c|x-y|^{\xi}$$

where $\zeta = l + \xi$ for some integer $l \ge 2m$ and $0 < \xi \le 1$.

Note that the assumption (A.1) implies that the p-th order cumulant spectrum (see Brillinger 1981, for definition) has bounded and uniformly continuous first derivative.

Theorem 1. Under the assumptions (A.1) with $p = 2, 4$ and (A.2) with $m = 1$, we have as $n \to \infty$ and $h \to 0$ with $nh \to \infty$,

$$MISE(h) = 2\pi n^{-1}h^{-1}R(fJ)R(K) + \frac{1}{4}h^4 R(f''J)\mu_2^2(K) + o(n^{-1}h^{-1} + h^4).$$

Here and below $J(x) = 1$ if $-\pi < x < \pi$ and 0 otherwise, $R(g) = \int g^2(x)dx$, $\mu_j(K) = \int x^j K(x)dx$. One useful feature of this representation is that it provides a simple version of an asymptotically optimal bandwidth. In particular, it is minimized by

$$h_0 = n^{-1/5}\{R(fJ)/R(f''J)\}^{1/5}\{2\pi R(K)/\mu_2^2(K)\}^{1/5}.$$

This motivates using the bandwidth

$$\hat{h} = n^{-1/5}\{\hat{R}(fJ)/\hat{R}(f''J)\}^{1/5}\{2\pi R(K)/\mu_2^2(K)\}^{1/5}, \qquad (2.3)$$

where $\hat{R}(fJ)$ and $\hat{R}(f''J)$ are some suitable estimtes of $R(fJ)$ and $R(f''J)$.

A natural estimate of $R(f^{(m)}J)$ is obtained by simply replacing $f(\lambda)$ by $\hat{f}_b(\lambda) = W_b * I_n(\lambda)$ where the bandwidth b and the kernel W are allowed to be different from h and K in (2.1). Hence we have $\hat{R}(f^{(m)}J) = R(\hat{f}_b^{(m)}J)$, and this is the type of estimates considered in Lee et al.(1993). Lee et al. give a specification of b^* which cancels two leading bias terms of $\hat{R}$, and hence gives the best mean squared error property. We briefly outline their results here. To do this, we need the following assumption on the kernel W.

(A.3) W is a symmetric kernel of order $2k$ ($k \ge 1$). i.e.,

$$\mu_0(W) = 1,\ \mu_1(W) = \cdots = \mu_{2k-1}(W) = 0,\ \mu_{2k}(W) \neq 0,$$

with a bounded support, and the m-th derivative $W^{(m)}$ satisfies the Lipschitz condition,

$$|W^{(m)}(x) - W^{m)}(y)| \leq c|x - y|$$

for all x and y, and for some positive constant c.

Let

$$c_1 = 2(-1)^k \mu_{2k}(W) R(f^{(m+k)}J)/(2k)!$$

$$c_2 = 2\pi R(W^{(m)}) R(fJ).$$

Lee et al. show that under (A.1) with $p = 2, \ldots, 8$, (A.2) and (A.3), the asymptotic bias and variance of $\hat{R}(f^{(m)}J)$ are given by, respectively, $c_1 b^{2k} + c_2 n^{-1} b^{-2m-1} + O(b^{2\nu})$ and $O(n^{-1} + n^{-2} b^{-4m-1})$, where $\nu = min\{(k+1), (\zeta - m)\}$. Now, if we use a kernel W with $(-1)^k \mu_{2k}(W) < 0$ so that $c_1 < 0$, then taking

$$b^* = (-c_1^{-1} c_2 n^{-1})^{1/(2m+2k+1)} \tag{2.4}$$

cancels the two leading bias terms, and so yields

$$\hat{R}(f^{(m)}J) - R(f^{(m)}J) = \begin{cases} O_p(n^{-(4k+1)/(4m+4k+2)}) & \text{if } k < m \\ O_p(n^{-1/2}) & \text{if } k \geq m \end{cases} \tag{2.5}$$

when $\zeta > k + m + \frac{1}{4}$.

Note that

$$\hat{h}/h_0 - 1 = O_p(\hat{R}(fJ) - R(fJ)) + O_p(\hat{R}(f''J) - R(f''J)). \tag{2.6}$$

From now on, write $\hat{h}(b_0, b_2)$ for $\hat{h}$ in (2.3) to stress its dependence on the bandwidths b_0 and b_2 used for $\hat{R}(fJ)$ and $\hat{R}(f''J)$ respectively. Suppose we use nonnegative kernels ($k = 1$) for $\hat{R}(fJ)$ and kernels of order k for $\hat{R}(f''J)$. Writing D_0 and D_2 for the constant factors $-c_1^{-1} c_2$ of the optimal bandwidths prescribed in (2.4) for $\hat{R}(fJ)$ and $\hat{R}(f''J)$,

$$b_0^* = (D_0 n^{-1})^{1/3} \quad , \qquad b_2^* = (D_2 n^{-1})^{1/(2k+5)}. \tag{2.7}$$

Then (2.5) and (2.6) together imply that if $\zeta > k + 2 + 1/4$, then

$$\hat{h}(b_0^*, b_2^*)/h_0 - 1 = \begin{cases} O_p(n^{-5/14}) & \text{if } k = 1 \\ O_p(n^{-1/2}) & \text{if } k \geq 2. \end{cases} \tag{2.8}$$

Of course, the formula (2.7) for b_0^* and b_2^* is not praticable because of their dependence on the unknown functionals $R(fJ)$, $R(f'J)$ and $R(f^{(k+2)}J)$ of f. Let $\hat{D}_0$ and $\hat{D}_2$ be some estimates of D_0 and D_2. The question is how accurate these estimates should be to ensure the same rate specified in (2.8). Note that the D's are chosen to cancel the two leading bias terms of $\hat{R}$, and a bit of mistuning for these constants may reintroduce non-negligible bias terms. For D_0, however, no precision of $\hat{D}_0$ is required since the two leading bias terms (of order $n^{-2/3}$ in this case) are already within $O(n^{-1/2})$. Putting aside practicability, this means that one can use any value for $\hat{D}_0$. A more practical approach is to replace f in D_0 with some reference spectral density. For example, the spectral density of MA(1) process may be used where the parameter is estimated by standard methods. For D_2, more than mere consistency is necessary. In this case, the reintroduced bias is of the same order as $n^{-2k/(2k+5)}(1 - D_2\hat{D}_2^{-1})$. Hence, whenever $\hat{D}_2 - D_2 = o_p(n^{-1/(4k+10)})$, (2.5) is valid.

Theorem 2. Under the assumptions (A.1) with $p = 2, ..., 8$, (A.2) and (A.3), if $\zeta > k + 2 + 1/4$ and $\hat{D}_2 - D_2 = o_p(n^{-1/(4k+10)})$, then

$$\hat{h}(\hat{b}_0, \hat{b}_2)/h_0 - 1 = \begin{cases} O_p(n^{-5/14}) & \text{if } k = 1 \\ O_p(n^{-1/2}) & \text{if } k \geq 2. \end{cases}$$

To get $\hat{D}_2$ satisfying the condition of Theorem 2, one may use $\hat{R}(f^{(m)}J)$ for various values of m. In this case, the estimates again need choices of bandwidths. But, since no great accuracy is required at this stage, one may use the optimal bandwidths prescribed in (2.4) with f in the unknown functionals of c_1 and c_2 replaced by some reference spectral density. This will be enough to afford the desired accuracy for $\hat{D}_2$.

3. A SIMULATION RESULT

The error criterion we use in this simulation study is

$$ISE(h) = \int_0^{\pi} [\hat{f}_h(\lambda) - f(\lambda)]^2 d\lambda, \tag{3.1}$$

Using the discretized version (2.2) of $\hat{f}_h(\lambda)$ and discretizing further the integral in the same way, (3.1) can be approximated by

$$ISE^D(h) = h^{-1}\delta^2 \sum_{j\in J} \sum_{l=1-M}^{M-1} K(\frac{l}{M}) I_n(\delta(j+l)) f(j\delta),$$

where $J = \{j : 0 \leq j \leq [\pi/\delta - 1/2] - 1\}$ and $[x]$ denotes the largest integer smaller than x. We use the quartic kernel, $K(x) = W(x) = (15/16)(1 - x^2)^2$ $I_{[-1,1]}(x)$ in all the estimation procedures. Note that this is a nonnegative kernel, and so is of order 2(k=1). the bandwidth $\hat{b}_0$ and $\hat{b}_2$ used for constructing $\hat{R}(fJ)$ and $\hat{R}(f''J)$ are:

$$\hat{b}_0 = [2\pi R(W) R(g_{\hat{\beta}} J)/\{\mu_2(W) R(g'_{\hat{\beta}} J)\}]^{1/3} n^{-1/3}$$

$$\hat{b}_2 = [2\pi R(W^{(2)}) \hat{R}_0/\{\mu_2(W)\hat{R}_3\}]^{1/7} n^{-1/7}, \tag{3.2}$$

where g_β denotes the spectral density of MA(1) process, $X(t) = \epsilon(t) - \beta\epsilon(t-1)$ with $\epsilon(t)$ the standard normal, $\hat{\beta}$ is the maximum likelihood estimate under this model. Here $R(fJ)$ and $R(f^{(3)}J)$ in b_2^* are estimated by $\hat{R}_0$ and $\hat{R}_3$:

$$\hat{R}_0 = R(\hat{f}_{\hat{b}_0} J), \; \hat{R}_3 = R(\hat{f}_{\hat{b}_3}^{(3)} J)$$

where

$$\hat{b}_3 = [2\pi R(W^{(3)}) R(g_{\hat{\beta}} J)/\{\mu_2(W) R(g_{\hat{\beta}}^{(4)} J)\}]^{1/9} n^{-1/9}.$$

All the estimates of $R(f^{(m)}J)$ used are the discretized version introduced in Lee et al. (1993). In fact,

$$\hat{R}(f^{(m)}J)$$
$$= \hat{b}_m^{-2(m+1)} \hat{\delta}_m^3 \sum_{j\in J_m} \sum_{l=1-M}^{M-1} \sum_{k=1-M}^{M-1} W^{(m)}(\frac{l}{M}) W^{(m)}(\frac{k}{M}) I_n(\hat{\delta}_m(j+l)) I_n(\hat{\delta}_m(j+k))$$

where $\hat{\delta}_m = \hat{b}_m/M$ and $J_m = \{j : -[\pi/\hat{\delta}_m - 1/2] - 1 \le j \le [\pi/\hat{\delta}_m - 1/2] + 1\}$.

The stationary time series models chosen for this study are

$$MA(1) : X(t) = \epsilon(t) - \beta\epsilon(t-1)$$
$$AR(1) : X(t) = \phi X(t-1) + \epsilon(t).$$

The values of β and ϕ selected are $\pm.3$ and $\pm.6$. The distribution of $\epsilon(t)$ is taken the standard normal, and $M = 30$ is used.

We compare $\hat{f}_{\hat{h}}$ and $\hat{f}_{h_0}$, and the means of comparison is the expected integrated squared error. Table 1 contains the Monte Carlo estimate of $E\{ISE^D(\hat{h})\}$ and $E\{ISE^D(h_0)\}$ based on 500 pseudo time series data of size 100. For taking into account the Monte Carlo variability, we computed the standard error of the Monte Carlo values, defined by $SE = \hat{\sigma}/\sqrt{500}$ where

$$\hat{\sigma}^2 = \sum_{i=1}^{500} (ISE_{(i)} - \hat{E}(ISE))^2/500.$$

For MA(1) processes with $\beta = \pm.3$, $\hat{f}_{\hat{h}}$ is significantly better than $\hat{f}_{h_0}$. This is not so surprising since h_0 is the minimizer of the asymptotic MISE, not ISE. For $\beta = \pm.6$, $\hat{f}_{\hat{h}}$ is still better but not so significantly. For AR(1) processes, the two show equivalent performance.

4. PROOF OF THEOREM 1

Using Lemma 2 of Lee et al.(1993)

$$\begin{aligned}\int_{-\pi}^{\pi} \{E\hat{f}_h(\lambda) - f(\lambda)\}^2 d\lambda &= \int_{-\pi}^{\pi} \{K_h * f(\lambda) + O(n^{-1}) - f(\lambda)\}^2 d\lambda \\ &= \frac{1}{4}\mu_2^2(K)R(f''J)h^4 + o(h^4) + O(n^{-1}) \\ &= \frac{1}{4}\mu_2^2(K)R(f''J)h^4 + o(h^4 + n^{-1}h^{-1}).\end{aligned}$$

By Lemma 1 and 2 of Lee et al. and defining $F_n(\alpha) = (2\pi n)^{-1}(sin\frac{n}{2}\alpha/sin\frac{\alpha}{2})^2$, we also have

$$
\begin{aligned}
&\int_{-\pi}^{\pi} Var\{\hat{f}_h(\lambda)\}d\lambda \\
&= \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} K_h * K_h(x)Cov\{I_n(\lambda - x), I_n(\lambda)\}d\lambda dx \\
&= \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} K_h * K_h(x)(2\pi n^{-1})\{F_n(x) + F_n(2\lambda - x)\}f^2(\lambda)d\lambda dx + O(n^{-1}) \\
&= 2\pi n^{-1}[K_h * K_h * F_n(0)R(fJ) + K_h * K_h * F_n * \{\frac{1}{2}f^2(\frac{\cdot}{2})J(\frac{\cdot}{2})\}(0)] + O(n^{-1}) \\
&= 2\pi n^{-1}h^{-1}\{K * K(0)R(fJ) + n^{-1}h^{-1}log(nh)\} + O(n^{-1}) \\
&= 2\pi n^{-1}h^{-1}R(K)R(fJ) + o(n^{-1}h^{-1}).
\end{aligned}
$$

Now the theorem follows.

REFERENCES

(1) Beltrão, K.I. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate, *Journal of Time Series Analysis*, 8, 21-38.

(2) Brillinger, D.R. (1981). *Time Series Data Analysis and Theory.* Holden-Day, San Francisco.

(3) Hurvich, C.M. (1985). Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods, *Journal of American Statistical Association*, 80, 933-940.

(4) Lee, Y.H., Cho, S., Kim, W.C. and Park, B.U. (1993). On estimating integrated squared spectral density derivatives, preprint.

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

(6) Robinson, P.M. (1991). Automatic frequency domain inference on semiparametric and nonparametric models, *Econometrica*, 59, 1329-1363.

Table 1. Based on 500 pseudo time series data of size 100. For each process the first line pertains to $\hat{h}$, the second h_0. EISE denotes the Monte Carlo estimate of $E\{ISE^D(\hat{h})\}$ or $E\{ISE^D(h_0)\}$, and SE its standard error as defined in the text.

		EISE	SE
	β= .3	.3118	.0035
		.3259	.0025
	β=-.3	.3227	.0039
MA(1)		.3839	.0036
	β= .6	.5691	.0077
		.5905	.0049
	β=-.6	.5891	.0071
		.6051	.0054
	ϕ= .3	1.080	.0043
		1.087	.0038
	ϕ=-.3	1.067	.0043
AR(1)		1.082	.0035
	ϕ= .6	2.197	.0185
		2.168	.0149
	ϕ=-.6	2.157	.0175
		2.146	.0133