A Simple Span Selector for Periodogram Smoothing

Thomas C. M. Lee

Biometrika, Vol. 84, No. 4 (Dec., 1997), 965-969.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199712%2984%3A4%3C965%3AASSSFP%3E2.0.CO%3B2-1

Biometrika is currently published by Biometrika Trust.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@umich.edu.

http://www.jstor.org/
Wed Mar 10 11:57:24 2004

Biometrika (1997), **84**, 4, *pp.* 965–969
Printed in Great Britain

A simple span selector for periodogram smoothing

By THOMAS C. M. LEE

Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637, U.S.A.
e-mail: tlee@galton.uchicago.edu

SUMMARY

One approach to estimating the spectral density of a stationary time series is to smooth the periodogram and one important component of this approach is the choice of the span for smoothing. This note proposes a new span selector which is based on unbiased risk estimation. The proposed span selector is simple, and does not impose strong conditions on the unknown spectrum. For example, it does not require the unknown spectrum to possess a second derivative, which is a typical requirement of most plug-in type or spline-based methods. The finite sample performance of the proposed span selector is illustrated via a small simulation.

Some key words: Bandwidth/span selection; Spectral density estimation; Unbiased risk estimation.

1. INTRODUCTION

Let $\{x_t\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ be a real-valued, zero mean stationary process with autocovariance function $\gamma_r = E(x_{t-r}x_t)$ $(r = 0, 1, \ldots)$ and spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_r \gamma_r \exp(-ir\omega), \quad \omega \in [0, 2\pi).$$

Suppose from one realisation of $\{x_t\}$ we observe $x_0, \ldots, x_{n-1}$ and compute the periodogram

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

Then the spectral density $f(\omega)$ can be estimated nonparametrically by applying weighted local smoothing to the periodogram. To simplify notation, let $\omega_j = 2\pi j/n$ and denote $I(\omega_j)$ and $f(\omega_j)$ by I_j and f_j, respectively, for $j = 0, \ldots, n-1$. The smoothed periodogram estimator $\hat{f}_j$ of f_j is

$$\hat{f}_j = \sum_{k=-p}^{p} w_{p,k} I_{j+k} \quad (j = 0, \ldots, n-1), \tag{1}$$

where $2p + 1$ is the span for smoothing and $w_{p,k}$ $(k = -p, \ldots, p)$ are nonnegative weights satisfying

$$w_{p,k} = w_{p,-k} \quad (k = 1, \ldots, p), \qquad \sum_{k=-p}^{p} w_{p,k} = 1. \tag{2}$$

The weight $w_{p,0}$ should also be a decreasing function of p. These conditions are satisfied by almost all suitably discretised kernel functions commonly used for smoothing. Since the spectral density is periodic with period 2π, periodic smoothing is used in (1) to handle boundary effects. That is, define $I_{-1} = I_{n-1}$, $I_n = I_0$ and so on. Fan (1992) demonstrated that weighted local linear regression is superior to the kernel type estimator (1) as it (i) adapts to different types of design, and (ii) does not require boundary modifications. However, in the present situation estimator (1) is expected to give similar performance as the data $(I_0, \ldots, I_{n-1})$ are equally spaced and boundary effects are handled by periodic smoothing. The parameter p plays the same role as the bandwidth in the nonparametric density and curve estimation contexts.

Here we propose an automatic method, based on unbiased risk estimation, for estimating the p which minimises the risk function

$$R(p)=\frac{1}{n}E\left\{\sum_{j=0}^{n-1}(f_j-\hat{f}_j)^2\right\}. \tag{3}$$

Two other widely-used and locally equivalent risk functions are

$$R'(p)=\frac{1}{n}E\left\{\sum_{j=0}^{n-1}(\log f_j-\log\hat{f}_j)^2\right\},\quad R''(p)=\frac{1}{n}E\left\{\sum_{j=0}^{n-1}\left(\frac{f_j-\hat{f}_j}{f_j}\right)^2\right\}. \tag{4}$$

However, with $R(p)$ the unknown spectral density f is not required to be strictly positive, i.e. zero values are allowed. Also, in the, as yet, unpublished paper 'Nonparametric estimation and simulation of two-dimensional Gaussian image textures' by T. C. M. Lee and M. Berman, it is shown that, in the context of image texture synthesis, $R(p)$ is the preferred risk function.

Nonparametric spectral density estimators other than (1) have been proposed in the literature. The windowed autocovariance estimator was one of the first. More recently, Wahba (1980) applied a cubic smoothing spline to smooth the log-periodogram, Chow & Grenander (1985) and Pawitan & O'Sullivan (1994) used a penalised Whittle likelihood approach, and Cameron (1987) developed some partitioning algorithms to approximate the spectral density by a step function. Hannan & Rissanen (1988) constructed a criterion based on the stochastic complexity for periodogram smoothing, Moulin (1994) applied wavelet thresholding techniques for log-periodogram smoothing, and Fan & Gijbels (1996, Ch. 6) describe unpublished work of J. Fan and E. Kreutzberger, in which they used local polynomial smoothers developed in Fan & Gijbels (1995) to smooth the periodogram and the log-periodogram. These authors also fitted the local Whittle likelihood for estimating the spectral density. Riedel & Sidorenko (1996) combined multiple tapering and plug-in techniques for adaptively smoothing the log-periodogram.

2. Unbiased risk estimation

2·1. *Background and cross-validation*

If all moments of x_t exist, the sum of all $|\gamma_r|$'s is bounded and n is large (Brillinger, 1981, Theorem 5.2.6), then

$$I_j \simeq f_j\varepsilon_j \quad (j=0,\ldots,n-1), \tag{5}$$

where ε_0, and $\varepsilon_{n/2}$ if n is even, are independent χ_1^2 random variables, and all other ε_j's are independent random variables distributed as the standard exponential distribution. If n is small, the data can be tapered in order to reduce the bias of the periodogram, and (5) remains approximately valid (Brillinger, 1981, Ch. 5). From now on, as in Moulin (1994) and Pawitan & O'Sullivan (1994), ε_0 and $\varepsilon_{n/2}$ will be treated as if they were standard exponential random variables and the approximation in (5) is assumed to be exact. The effect of these changes is asymptotically negligible. Hence we have the following model:

$$I_j=f_j\varepsilon_j \quad (j=0,\ldots,n-1), \tag{6}$$

where ε_j are independent standard exponential random variables. Note that $E(\varepsilon_j)=\mathrm{var}(\varepsilon_j)=1$ and $E(\varepsilon_j^2)=2$.

Since $R(p)$ is unknown, a natural approach to finding its minimiser is to form an unbiased estimator of $R(p)$ and choose p to minimise it. In a cross-validation method, which is presented in an unpublished report of F. Palmer and is mentioned in Hurvich (1985), p is chosen to be the minimiser p_{CV} of the cross-validation score

$$\mathrm{CV}(p)=\frac{1}{n}\sum_{j=1}^{n}(I_j-\hat{f}_{-i})^2=\frac{\mathrm{RSS}}{n(1-w_{p,0})^2}, \tag{7}$$

where $\hat{f}_{-i}$ is the usual 'leave-one-out' estimate for f_i and $\text{RSS} = \sum (I_j - \hat{f}_j)^2$ is the residual sum of squares. The last equality of (7) can be obtained by using equation (3) of Hurvich (1985). It is straightforward to show that $\text{CV}(p)$ is biased for $R(p)$ when n is finite. Since, for low-dimensional parameter estimation problems, biased estimators can lead to serious estimation errors, we seek an unbiased estimator for $R(p)$.

2·2. *The proposed span selector*

We begin by calculating $E(\text{RSS}) = \sum E(I_j^2 - 2I_j\hat{f}_j + \hat{f}_j^2)$. Since the ε_j are independent standard exponentials, then $E(I_j) = f_j$, $E(I_j^2) = E(f_j^2\varepsilon_j^2) = 2f_j^2$ and

$$E(I_j\hat{f}_j) = E\left(f_j\varepsilon_j \sum_{k=-p}^{p} w_{p,k} f_{j+k}\varepsilon_{j+k}\right) = E\left(w_{p,0} f_j^2 \varepsilon_j^2 + \sum_{k\neq 0} w_{p,k} f_j \varepsilon_j f_{j+k} \varepsilon_{j+k}\right)$$

$$= 2w_{p,0} f_j^2 + f_j \sum_{k\neq 0} w_{p,k} f_{j+k} = w_{p,0} f_j^2 + f_j \sum_{k=-p}^{p} w_{p,k} f_{j+k} = w_{p,0} f_j^2 + f_j E(\hat{f}_j).$$

Therefore,

$$E(I_j - \hat{f}_j)^2 = 2f_j^2 - 2\{w_{p,0} f_j^2 + f_j E(\hat{f}_j)\} + E(\hat{f}_j^2) = E(f_j - \hat{f}_j)^2 + (1 - 2w_{p,0}) f_j^2,$$

$$E(\text{RSS}) = nR(p) + (1 - 2w_{p,0}) E\left(\sum \frac{1}{2} I_j^2\right).$$

Thus

$$\hat{R}(p) = \frac{\text{RSS}}{n} - \frac{(1 - 2w_{p,0})}{2n} \sum I_j^2 \tag{8}$$

is an unbiased estimator of $R(p)$. We propose to choose the minimiser p_{ER} of $\hat{R}(p)$ as the span.

This idea of unbiased risk estimation has been used for bandwidth selection in the nonparametric curve estimation context (Rice, 1984; Chiu, 1990) and has its origin in Mallows' C_p (Mallows, 1973), but it has not been applied to nonparametric spectral estimation. The crucial difference in the present context is that the variance of the noise variables ε_j is known.

2·3. *Log-periodogram smoothing*

Unbiased risk estimation can also be applied to choose the span for log-periodogram smoothing. In this case it is natural to use $R'(p)$ defined by (4), and the first step is to transform the multiplicative model (6) into an additive model by taking a logarithmic transform:

$$y_j = \log I_j + \psi(1) = \log f_j + \xi_j \quad (j = 0, \ldots, n-1),$$

where ξ_j are independent zero mean random variables with variance $\pi^2/6$, and $\psi(x)$ is the digamma function with $\psi(1) = 0{\cdot}57722$; see Wahba (1980), in which a cubic smoothing spline was applied to smooth the y_j's, with smoothing parameter chosen by generalised cross-validation. Using the same technique as before, one can show that

$$\hat{R}'(p) = \frac{1}{n} \sum_{j=0}^{n-1} (y_j - \hat{g}_j)^2 - \frac{(1 - 2w_{p,0})\pi^2}{6n}$$

is an unbiased estimator of $R'(p)$. Here $\hat{g}_j = \sum w_{p,k} y_{j+k}$ is the estimate of $\log f_j$. Thus one can choose the minimiser of $\hat{R}'(p)$ as the span for log-periodogram smoothing.

3. SIMULATION

A simulation was conducted to evaluate the relative finite sample performance of the proposed span selector p_{ER} and the cross-validation based selector p_{CV}. The weights $w_{p,k}$'s in (1) used for this simulation study were

$$w_{p,k} = w'_{p,k}\left(\sum_{k=-p}^{p} w'_{p,k}\right)^{-1}, \quad w'_{p,k} = \frac{3}{4p}\left(1 - \frac{k^2}{p^2}\right) \quad (k = -p, \ldots, p).$$

These weights are a discrete version of the optimal kernel of order (0, 2) derived in Gasser, Müller & Mammitzsch (1985). They minimise the asymptotic optimal risk, and it is easy to verify that they satisfy conditions (2).

Four different test examples from the ARMA (α, β) model

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

were used: Example 1 was AR (3) with $a_1 = -1{\cdot}5$, $a_2 = 0{\cdot}7$ and $a_3 = -0{\cdot}1$; Example 2 was MA (4) with $b_1 = -0{\cdot}3$, $b_2 = -0{\cdot}6$, $b_3 = -0{\cdot}3$ and $b_4 = 0{\cdot}6$; Example 3 was AR (3) with $a_1 = 0{\cdot}9$, $a_2 = 0{\cdot}8$ and $a_3 = 0{\cdot}6$; and Example 4 was MA (3) with $b_1 = 0{\cdot}9$, $b_2 = 0{\cdot}8$ and $b_3 = 0{\cdot}6$. Examples 1 and 2 were used by Fan & Gijbels (1996, Ch. 6), Pawitan & O'Sullivan (1994) and Wahba (1980), and Examples 3 and 4 were used by Hurvich (1985).

For each example 200 independent series were simulated with $n = 256$, and for each series we computed the optimal risk $R_{\mathrm{opt}} = \min_p R(p)$, and the risks $R(p_{\mathrm{CV}})$ and $R(p_{\mathrm{ER}})$ using (3). We also computed the two risk-ratios $r_{\mathrm{CV}} = R(p_{\mathrm{CV}})/R_{\mathrm{opt}}$ and $r_{\mathrm{ER}} = R(p_{\mathrm{ER}})/R_{\mathrm{opt}}$. The values of the average risk-ratio differences $(r_{\mathrm{CV}} - r_{\mathrm{ER}})$ for Examples 1 to 4 respectively were $-0{\cdot}0083$, $0{\cdot}1505$, $0{\cdot}1533$ and $1{\cdot}6307$, with respective estimated standard deviations $0{\cdot}0257$, $0{\cdot}1082$, $0{\cdot}0393$ and $0{\cdot}4296$: the 200 simulations were insufficient to discriminate statistically between the two selectors for Examples 1 and 2, while for Examples 3 and 4 there is strong evidence that p_{ER} outperformed p_{CV}.

ACKNOWLEDGEMENT

This work is part of the author's PhD thesis. He would like to express his gratitude to his supervisors, Mark Berman and Vic Solo, for their continuous encouragement, guidance and patience. He would also like to thank a referee, an associate editor and the editor for valuable comments, which have led to this more succinct version of the paper.

REFERENCES

BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*. New York: Holden-Day.

CAMERON, M. A. (1987). An automatic non-parametric spectrum estimator. *J. Time Ser. Anal.* **8**, 379–87.

CHIU, S.-T. (1990). Why bandwidth selectors tend to choose smaller bandwidths, and a remedy. *Biometrika* **77**, 222–6.

CHOW, Y.-S. & GRENANDER, U. (1985). A sieve method for the spectral density. *Ann. Statist.* **13**, 998–1010.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Am. Statist. Assoc.* **87**, 998–1004.

FAN, J. & GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Statist. Soc.* B **57**, 371–94.

FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

GASSER, T., MÜLLER, H.-G. & MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. R. Statist. Soc.* B **47**, 238–52.

HANNAN, E. & RISSANEN, J. (1988). The width of a spectral window. *J. Appl. Prob.* **25**, 119–25.

HURVICH, C. M. (1985). Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *J. Am. Statist. Assoc.* **80**, 933–40.

MALLOWS, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–75.

MOULIN, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Sig. Process.* **42**, 3126–36.

PAWITAN, Y. & O'SULLIVAN, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Am. Statist. Assoc.* **89**, 600–10.

RICE, J. A. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215–30.
RIEDEL, K. S. & SIDORENKO, A. (1996). Adaptive smoothing of the log-spectrum with multiple tapering. *IEEE Trans. Sig. Process.* **14**, 1794–800.
WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc.* **75**, 122–32.

[*Received February* 1996. *Revised June* 1997]