OPTIMAL SPECTRAL KERNEL FOR LONG-RANGE DEPENDENT TIME SERIES

Miguel A. Delgado and Peter M. Robinson*

Abstract

We derive an optimal kernel $K(\lambda)$ for spectral averaging in the neighbourhood of a spectral peak corresponding to long range dependence. Unusually, $K(\lambda) \rightarrow 0$ as $\lambda \rightarrow 0$.

Key Words
Long Range Dependence; Spectral Analysis; Optimal Kernel.

1. Introduction

Let X_t, $t = 0, \pm 1, \pm 2, \ldots$, be a covariance stationary time series with lag-j autocovariance $\gamma_j = E\left[(X_j - E(X_0))(X_0 - E(X_0))\right]$ and spectral density $f(\lambda)$ satisfying $\gamma_j = \int_{-\pi}^{\pi} f(\lambda) \cos(j\lambda) \, d\lambda$, $j = 0, 1, \ldots$. It is assumed that X_t has long range dependence, in the sense that

$$f(\lambda) \sim L(1/\lambda) \, \lambda^{2H-1} \text{ as } \lambda \longrightarrow 0+, \tag{1.1}$$

for $1/2 < H < 1$, where $L(\lambda)$ is slowly varying at infinity. In order to study the behaviour of f at zero frequency, Robinson (1994 a, b) considered the averaged periodogram statistic

$$\hat{F}(\lambda_m) = \frac{2\pi}{n} \sum_{j=1}^{m} I(\lambda_j), \tag{1.2}$$

where $I(\lambda) = (2\pi n)^{-1} \left| \sum_{t=1}^{n} X_t \, e^{it\lambda} \right|^2$, $\lambda_j = 2\pi j/n$, and $1 \leq m < n/2$. The $I(\lambda_j)$ for $1 \leq j < n$ are invariant to location-shift in X_t, so no explicit mean correction is needed. The bandwidth number m is supposed to increase with n, but more slowly.

When $f(\lambda)$ is smooth at $\lambda = 0$, $\hat{F}(\lambda_m)/\lambda_m$ is a standard nonparametric estimate of $f(0)$ (see e.g. Brillinger 1975). However, Robinson (1994 a) demonstrated the usefulness of $\hat{F}(.)$ in case of (1.1). It can be used to consistently estimate H, even in the presence of $L(.)$ of unknown form; it, and an analogous averaged cross periodogram, can be used to consistently estimate the coefficient in a stationary long-range dependent co-integrated system; it can be used in consistent estimation of the limiting variance matrix of least squares estimates of regression coefficients in the presence of long-range dependent errors and certain types of regressor; it can be used to construct estimates with the same limiting distribution as efficient generalized least squares estimates of regression coefficients in the presence of long range dependent errors. In addition, the behaviour of (1.2) is of interest in case $\hat{F}(\lambda_m)$ is computed in the incorrect belief that $f(0) < \infty$.

1

As in the smooth spectrum case, and other problems of nonparametric estimation, the choice of bandwidth number m is important. Robinson (1994 b) derived an asymptotic approximation for the mean squared error

$$\stackrel{\wedge}{\text{MSE}} = \text{E}\left(\frac{\hat{F}(\lambda_m) - G(\lambda_m)}{G(\lambda_m)} \right)^2,$$ (1.3)

where

$$G(\lambda) \stackrel{\text{def}}{=} L\left(\frac{1}{\lambda}\right) \frac{\lambda^{2-2H}}{2-2H} \cong \int_0^\lambda L\left(\frac{1}{\omega}\right) \omega^{1-2H} \, d\omega,$$

and also derived formulae for m which minimize this approximation. We can regard (1.3) as an analogue of the usual mean squared error criterion

$$\text{E}\left(\frac{\hat{f}(0) - f(0)}{f(0)} \right)^2,$$ (1.4)

employed in case $0 < f(0) < \infty$, where $\hat{f}(0) = \hat{F}(\lambda_m)/\lambda_m$. Taking $L(1/\lambda) \equiv f(0)$ and $H = 1/2$ in (1.3) gives (1.4). Delgado and Robinson (1993) have further analyzed the approximation of (1.3), carried out numerical calculations, and provided feasible plug-in-versions of the optimal m.

In (1.2) equal weights of the $I(\lambda_j)$ are employed, but a weighted average of periodograms is often considered in case of estimation of a smooth spectral density, in both the theoretical literature and in practice (see e.g. Brillinger 1975). Indeed, bearing in mind the asymptotic equivalence of averaged periodogram spectrum estimates and the weighted autocovariance spectrum estimates stressed in much of the earlier literature, there seems to have been much greater stress on choice of weighting function than on choice of bandwidth, as reflected for example by many texts on spectral analysis (e.g. Blackman and Tukey 1959; Grenander and Rosenblatt 1957; Brillinger 1975). This contrasts with the problem of nonparametric probability density estimation and regression estimation, where interest has centred on the bandwidth problem. A likely reason for the concern with weighting in spectral analysis is the apparent great peakedness of many spectra, which causes difficulties in estimation both near peaks and, owing to leakage, at distant

2

frequencies. The same degree of peakedness is not commonly encountered in the study of probability density and regression functions.

A long-range dependent time series has an especially peaked spectrum, and so it seems worthwhile to investigate the optimal choice of weighting in estimating spectral mass near the peak. Section 2 of the paper considers a weighted generalization of (1.2), indicates the form of the optimal minimum mean squared error bandwidth for this, and derives the weights which minimize the bandwidth-optimized mean squared error. Interestingly, the optimal weights tend to zero at the frequency of the spectral peak and thus differ radically from optimal weights in smooth nonparametric problems.

2. Optimal weights

Let $K(\lambda)$ be a real function, called a spectral window or kernel, such that

$$K(\lambda) = K(-\lambda), \quad \int_{-\infty}^{\infty} |K(\omega)| \ d \omega < \infty. \tag{2.1}$$

For the time being we specify no normalization for $K(.)$. Define the coefficients

$$c_\theta = \int_{-\infty}^{\infty} |\omega|^\theta K(\omega) \ d\omega, \quad d_\phi = \int_{-\infty}^{\infty} |\omega|^\phi K(\omega)^2 \ d\omega,$$

taking it for granted that the integrals exist for relevant θ and ϕ values.

For m as defined previously, introduce

$$\hat{F}_K(\lambda_m) = \frac{2\pi\lambda_m}{n} \sum_{j=1}^{n-1} K_n(\lambda_j) \ I(\lambda_j), \tag{2.2}$$

where

$$K_n(\lambda) = \frac{n}{2m} \sum_{j=-\infty}^{\infty} K\left(\frac{n}{2m} (\lambda + 2\pi j)\right).$$

Notice that the Daniell (1948) kernel

3

$$K(\lambda) = \begin{cases} (2\pi)^{-1}, & |\lambda| \leq \pi, \\ 0, & |\lambda| > \pi, \end{cases} \tag{2.3}$$

leads to $\hat{F}_K(\lambda_m) = \hat{F}(\lambda_m)$, so that (2.2) generalizes (1.2). In fact $\hat{F}_K(\lambda_m)$ can be thought of as estimating

$$G_K(\lambda_m) = \frac{2 \, (1 - H)}{\pi^{1-2H}} \, G(\lambda_m) \, c_{1-2H}$$

$$= L\left(\frac{1}{\lambda}\right) \, \lambda_m^{2-2H} \, \frac{c_{1-2H}}{\pi^{1-2H}},$$

with the latter reducing to $G(\lambda_m)$ under (2.3).

Now consider the scaled mean squared error

$$\hat{MSE}_K = E\left(\frac{\hat{F}_K(\lambda_m) - G_K(\lambda_m)}{G_K(\lambda_m)} \right)^2,$$

which generalizes (1.3). For simplicity let X_t be Gaussian (though the results derived below hold more generally, cf Robinson (1994 b)) and strengthen (1.1) to the assumption

$$f(\lambda) = L\left(\frac{1}{\lambda}\right) \, \lambda_m^{1-2H} \left(1 + E_{\alpha H} \, \lambda^\alpha + o(\lambda^\alpha)\right) \text{ as } \lambda \longrightarrow 0+, \tag{2.4}$$

for some $H \in (1/2, 3/4)$, some $\alpha \in (0, 2]$ and some $E_{\alpha H} \in (-\infty, \infty)$. We also assume that $f(\lambda)$ is differentiable in a neighbourhood $(0, \varepsilon)$ for some $\varepsilon > 0$, and

$$\frac{df(\lambda)}{d\lambda} = o\left(L\left(\frac{1}{\lambda}\right) \, \lambda^{-2H} \right), \text{ as } \lambda \longrightarrow 0. \tag{2.5}$$

Assumption (2.4) is an extension of the usual smoothness condition imposed in smooth spectral analysis, and for $L(\lambda) \equiv C \in (0, \infty)$, and $\alpha = 2$, both (2.4)

4

and (2.5) are satisfied by two familiar models for long-range dependence, the "fractional noise" process with spectrum

$$f(\lambda) = \frac{\sigma^2 \sin(\pi H)}{\pi} \, \Gamma(2H - 1) \, (1 - \cos \lambda) \sum_{j=-\infty}^{\infty} \left| \lambda + 2\pi j \right|^{-2\pi-1},$$

and the "fractional ARIMA" process with spectrum

$$f(\lambda) = \frac{\sigma^2}{2\pi} \, \left| 1 - e^{i\lambda} \right|^{1-2H} \, \frac{\left| b(e^{i\lambda}) \right|^2}{\left| a(e^{i\lambda}) \right|^2},$$

where $\sigma^2 > 0$, and a and b are polynomials of finite degree having no roots in or on the unit circle.

Under conditions (2.4) and (2.5), and also (2.1) and

$$\frac{1}{m} + \frac{m}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty, \tag{2.6}$$

a straightforward extension of results in Robinson (1994 b) indicates that

$$\widehat{MSE}_K = \frac{d_{2-4H}}{c_{1-2H}^2} \, \frac{2\pi}{m} + \left(\frac{c_{1-2H+\alpha}}{c_{1-2H}} \, \frac{E_{\alpha H}}{\pi^\alpha} \right)^2 \lambda_m^{2\alpha} \text{ as } n \longrightarrow \infty. \tag{2.7}$$

Notice that $\widehat{MSE}_k$ is free of the slowly varying function $L(.)$, which can be of unknown form. (2.7) reduces to Robinson's (1994 b) result

$$\widehat{MSE} = 4(1-H)^2 \left[\frac{1}{(3-4H)\,m} + \left\{ \frac{E_{\alpha H}}{2 - 2H + \alpha} \right\}^2 \lambda_m^{2\alpha} \right], \text{ as } n \longrightarrow \infty,$$

in case $K(.)$ is given by (2.3).

The property (2.6) implies that $\hat{F}_K(\lambda_m)/G_K(\lambda_m) \xrightarrow{P} 1$, and thence that $\hat{F}_K(\lambda_m)$ can be useful in the ways ascribed to $\hat{F}(\lambda_m)$. Moreover, the right side of (2.7) can be minimized with respect to m by

5

$$\hat{m}_K(H) = \left(\frac{2\pi d_{2-4H}}{2\alpha 2^{2\alpha} E_{\alpha H}^2 c_{1-2H+\alpha}^2} \right)^{1/(2\alpha+1)} n^{2\alpha/(2\alpha+1)}. \tag{2.8}$$

Thus, choice of K does not affect the rate of convergence of the optimal m, but it does affect its scale factor. We can substitute (2.8) for m in (2.7) and then minimize again with respect to K. We accomplish this by the calculus of variations. We introduce the constraints

$$c_{1-2H} = \frac{\pi^{1-2H}}{2-2H}, \quad c_{1-2H+\alpha} = \frac{\pi^{1-2H+\alpha}}{2-2H+\alpha}, \tag{2.9}$$

which apply directly to the Daniell kernel (2.3). The Lagrangean is

$$d_{2-4H} - 2\,\xi_1 \left(c_{1-2H} - \frac{\pi^{1-2H}}{2-2H} \right) + 2\,\xi_2 \left(c_{1-2H+\alpha} - \frac{\pi^{1-2H+\alpha}}{2-2H+\alpha} \right),$$

so, denoting by $\Delta K(\lambda)$ a small increment of K, we need

$$\int_0^\infty \lambda^{2-4H}\, K(\lambda)\, \Delta K(\lambda)\, d\lambda - \xi_1 \int_0^\infty \lambda^{1-2H}\, \Delta K(\lambda)\, d\lambda + \xi_2 \int_0^\infty \lambda^{1-2H+\alpha}\, \Delta K(\lambda)\, d\lambda = 0,$$

that is

$$\lambda^{2-4H}\, K(\lambda) - \xi_1\, \lambda^{1-2H} + \xi_2\, \lambda^{1-2H+\alpha} = 0.$$

Thus we take

$$K(\lambda) = \begin{cases} \xi_1\, |\lambda|^{2H-1} - \xi_2\, |\lambda|^{2H-1+\alpha}, & |\lambda| \leq (\xi_1/\xi_2)^{1/\alpha} \\ 0 & , \quad |\lambda| > (\xi_1/\xi_2)^{1/\alpha}. \end{cases} \tag{2.10}$$

We deduce that

$$\xi_1 = \frac{(\alpha + 1)(2 - 2H + \alpha)^{1/\alpha}}{2\pi^{2H} \alpha (2\alpha + 1)^{1/\alpha} (2 - 2H)^{1+1/\alpha}}, \quad \xi_2 = \frac{\xi_1 (2 - 2H + \alpha)}{\pi^{\alpha} (2\alpha + 1)(2 - 2H)}.$$

In the short memory case, corresponding to H = 1/2, the optimal kernel (2.10) corresponds to the familiar kernel of Bartlett (1963); because the latter is calculated with the constraint $c_0 = 1$, instead of the first part of (2.9), our version is flatter. For H>1/2, $K(\lambda)$ is bimodal and $K(\lambda) \longrightarrow 0$ as $\lambda \longrightarrow 0$, so that the contribution from the very lowest frequencies is being downweighted. This is consistent with the suggestion of Künsch (1986) that, in semiparametric estimation of H, the periodogram at such frequencies be trimmed out due to their anomalous behaviour in case of long range dependence.

Figure 1 plots the optimal K (2.10) for $\alpha = 2$ (the leading choice of α in the smooth spectral analysis literature) with H= 0.6 and 0.7, along with, for the sake of comparison, the Daniell and Barltett kernels which are often used in spectral analysis.

FIGURE 1 ABOUT HERE

For general $\alpha \in (0, 2]$, the scaled mean squared error evaluated at $\hat{m}_K = \hat{m}_K(H)$ is proportional to

$$\frac{d_{2-4H}^{\frac{2\alpha}{2\alpha+1}} c_{1-2H+\alpha}^{\frac{2}{2\alpha+1}}}{c_{1-2H}^2}, \tag{2.11}$$

where the proportionality constant is identical for all kernels. For the optimal kernel (2.10),

$$d_{2-4H} = 2 \left[1 - \frac{2}{\alpha + 1} + \frac{1}{2\alpha + 1} \right] \xi_1^2 \, (\xi_1/\xi_2)^{1/\alpha}$$

7

$$= \frac{(\alpha^2 - \alpha - 1)\ (\alpha + 1)\ (2 - 2H + \alpha)^{1/\alpha}}{\pi^{4H-1}\ \alpha^2\ (2\alpha + 1)^{1+1/\alpha}\ (2 - 2H)^{2+1/\alpha}}\ . \tag{2.12}$$

For the Daniell kernel

$$d_{2-4H} = \frac{1}{\pi^{4H-1}\ 2(3-4H)}\ . \tag{2.13}$$

Figure 2 shows the relative efficiency of the Daniell kernel, calculated as the ratio of (2.11) for the Daniell kernel to (2.11) for the optimal kernel (2.10) with $\alpha=2$; this is $\{(2.12)/(2.13)\}^{2\alpha/(2\alpha+1)}$, in view of (2.8). The Daniell kernel is satisfactory for the smallest values of H, but deteroriates significantly for larger H.

FIGURE 2 ABOUT HERE

We have also performed a small Monte Carlo experiment in order to compare MSE of estimates of $G(\lambda_{\hat{m}_K})$, based on the Daniell and the optimal kernels. We report, in Table 1, Monte Carlo efficiency for the Daniell kernel relative to the optimal kernel for different values of H. We have generated data according to a fractional Gaussian autoregressive process with spectrum

$$f(\lambda) = \frac{1}{2\pi}\ \left| 1 - e^{i\lambda} \right|^{1-2H}\ \frac{1}{\left| 1 - a\ e^{i\lambda} \right|^2}\ ,$$

with $a = 0.5$. We provide results for H= 0.6, 0.65 and 0.7. The Daniell kernel's efficiency falls off as H increases, and the optimal kernel is clearly superior when H= 0.7. For each H value, there is good stability across sample size.

8

TABLE 1

Monte Carlo efficiencies of Daniell versus Optimal kernels for several H
values and sample sizes, based on 5,000 replications.
(MSE for Daniell kernel in parenthesis).

	n= 250	n= 500	n= 1000
H= 0.60	0.7587	0.7167	0.7091
	(0.0658)	(0.0420)	(0.0251)
H= 0.65	0.6239	0.5613	0.5706
	(0.0778)	(0.0518)	(0.0330)
H= 0.70	0.4832	0.4003	0.3113
	(0.1036)	(0.0742)	(0.0514)

REFERENCES

Bartlett, M.S. (1963), Statistical estimation of density functions, *Sankhya Ser. A* **25**, 245–254.

Blackman, R.B. and J.W. Tukey (1959), *The Measurement of Power Spectra*, Addison–Wesley, Reading, Mass.

Brillinger, D.R. (1975), *Time Series, Data Analysis and Theory*. Holden–Day, San Francisco.

Daniell, P.J. (1945), Discussion on "Symposium on spectral approach to time series", *J. Roy. Statist. Soc. Suppl.* **8**, 88–90.

Delgado, M.A. and P.M. Robinson (1993), Optimal spectral bandwidth for long memory, Working Paper 93–34, Universidad Carlos III de Madrid.

Grenander, U. and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

Künsch, H.R. (1986), Discrimination between monotonic trends and long range dependence, *J. Appl. Probab.* **23**, 1025–1030.

Robinson, P.M. (1994 a), Semiparametric analysis of long–memory time series, *Ann. Statist.* **22**, 515–539.

Robinson, P.M. (1994 b), Rates of convergence and optimal spectral bandwidth for long range dependence, *Probab. Theory Rel. Fields.* **99**, 443–473.
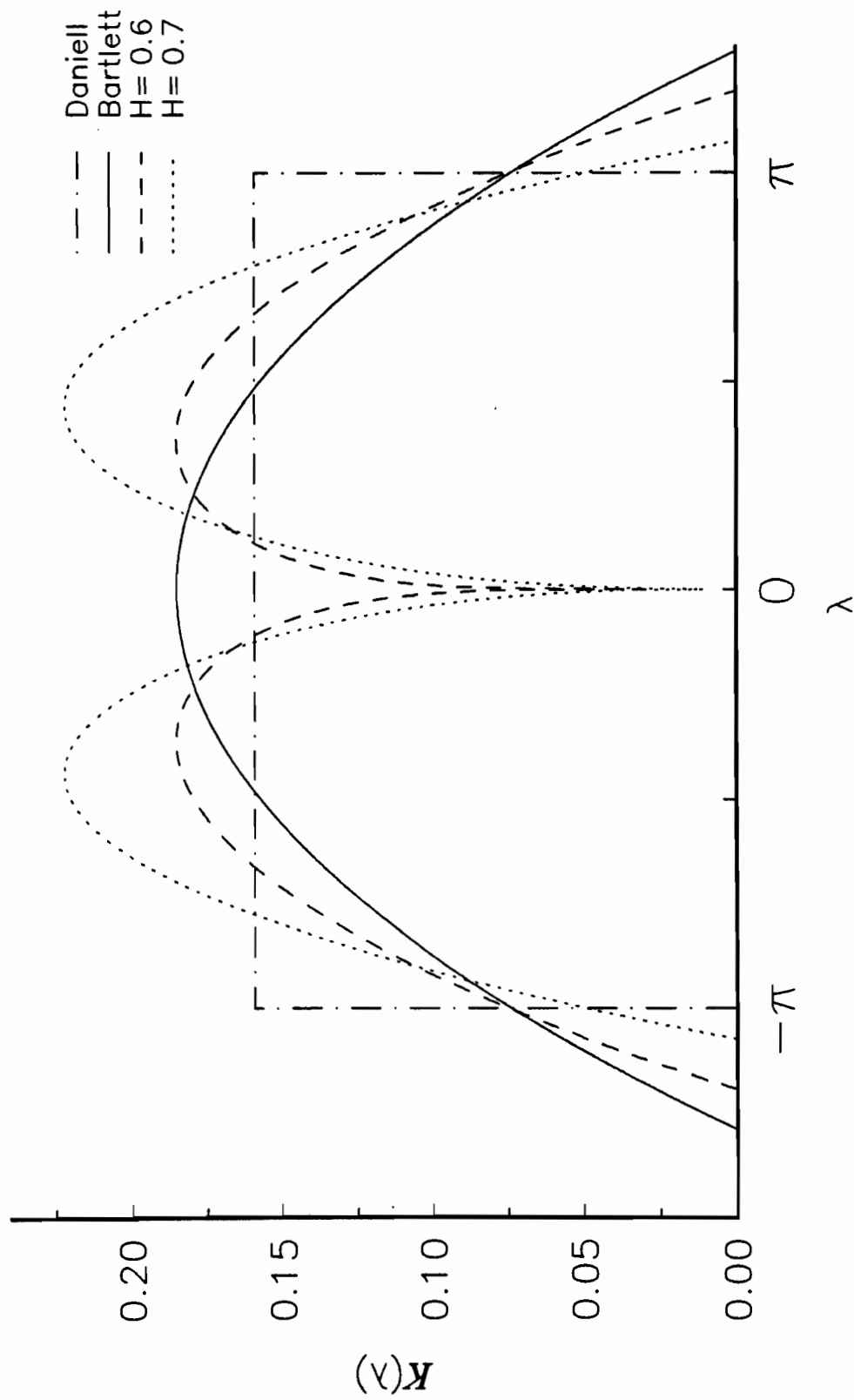
FIGURE 1

Optimal kernel functions for different H values

FIGURE 2

Relative efficiency with respect to Daniell's kernel