Statistica Sinica **21** (2011), 1115-1144

SPECTRAL DENSITY ESTIMATION THROUGH A REGULARIZED INVERSE PROBLEM

Chunfeng Huang, Tailen Hsing and Noel Cressie

Indiana University, University of Michigan and Ohio State University

Abstract: In the study of stationary stochastic processes on the real line, the covariance function and the spectral density function are parameters of considerable interest. They are equivalent ways of expressing the temporal dependence in the process. In this article, we consider the spectral density function and propose a new estimator that is not based on the periodogram; the estimator is derived through a regularized inverse problem. A further feature of the estimator is that the data are not required to be observed on a grid. When the regularization condition is based on the function's first derivative, we give the estimator in closed form as well as a bound on its mean squared error. Our numerical studies compare our new estimator of the spectral density to several well known estimators, and we demonstrate its increased statistical efficiency and much faster computation time.

Key words and phrases: Non-gridded data, periodogram, Sobolev space, stationary process, time series.

1. Introduction

Consider a real-valued second-order stationary process $X = \{X(t), t \in \mathbb{R}\}$ with mean 0 and covariance function R. Assume that X has a spectral density function f so that

$$R(t) = \int e^{it\omega} f(\omega) d\omega = 2 \int_0^\infty \cos(t\omega) f(\omega) d\omega, \tag{1.1}$$

$$f(\omega) = (2\pi)^{-1} \int e^{-i\omega t} R(t) dt. \tag{1.2}$$

The nonparametric estimation of f is a classical problem in the study of stochastic processes. Suppose that complete data $\{X(t), 0 \le t \le T\}$ are available. Define the periodogram,

$$I_T(\omega) = T^{-1} \Big| \int_0^T X(t) e^{-it\omega} dt \Big|^2 = \int_{-T}^T e^{-it\omega} \hat{R}(t) dt,$$

where $\hat{R}(t) = T^{-1} \int_0^{T-|t|} X(u) X(u + |t|) du$. From (1.2), $I_T(\omega)$ is asymptotically unbiased for $2\pi f(\omega)$, as $T \to \infty$. However, $I_T(\omega)$ is not consistent in that the

variance does not tend to zero as $T \to \infty$. From the references at the end of this paragraph, consistent estimators of f can be obtained by local averaging of the periodogram. Much of the spectral-analysis literature has focused on how to make this procedure work using various smoothing techniques. In that regard, discrete-index stochastic processes (i.e., time series) have received more attention than continuous-index stochastic processes (such as represented in (1.1) and (1.2)). For early literature, see Bartlett (1950), Grenander and Rosenblatt (1953), Jenkins and Watts (1968), Parzen (1957), and Priestley (1981). More recent works on data-driven procedures based on the periodogram include Beltrão and Bloomfield (1987), Fan and Kreutzberger (1998), Hurvich (1985), Hurvich and Beltrão (1990), and Pawitan and O'Sullivan (1994), to name a few.

In this paper we consider an approach for estimating f without directly using (1.2), and indeed without using the periodogram. The basic idea of our approach is to estimate f from an estimate of R by solving a regularized inverse problem; that is, "algorithmic inversion" takes the place of analytic inversion. The potential of our general approach can be realized in a number of settings, including stationary and intrinsically stationary processes indexed in $\mathbb{R}^k$ (cf., Cressie (1993), Matheron (1973), and Yaglom (1987)), each of which entails considerations unique to that setting. In this paper, we focus on stationary processes indexed by $\mathbb{R}$. Clearly, the data will be collected on a finite subset set of $\mathbb{R}$. In Section 2, we give the notation and describe the basic methodology that leads to the new spectral density estimator. In Section 3, we consider gridded data (i.e., time series data) and derive a representation of the estimator for that case. We describe a weighted cross-validation approach for choosing the smoothing parameter in our method, and we compare that procedure with a number of benchmark spectral procedures for stationary time series. In addition, Section 3 considers the computation of bounds on the mean squared error of the estimator. In Section 4, we demonstrate how our method can be modified to accommodate non-gridded (i.e., irregularly located) data. A mean squared error bound of the estimator is also obtained, and a comparison is made with existing procedures in the literature. Discussion and conclusions are given in Section 5. All of the proofs are given in the Appendix.

2. Methodology

In the developments below, it is convenient notationally to absorb the constant 2 into f in equation (1.1); that is, from now on we write

$$R(t) = \int_0^\infty \cos(t\omega) f(\omega) d\omega, \tag{2.1}$$

and keep in mind that we are estimating twice the usual spectral density function. Suppose that we observe the process $X(t)$ at $t = t_i, 1 \le i \le N$. Since X has

mean zero, the product $X(t_i)X(t_j)$ is an unbiased estimator of $R(t_i - t_j) = \int_0^\infty \cos((t_i - t_j)\omega) f(\omega) d\omega$. Thus, intuitively, the following sum of squares is small for a function g close to f:

$$\sum_{1 \le i,j \le N} \Big[X(t_i)X(t_j) - \int_0^\infty \cos((t_i - t_j)\omega) g(\omega) d\omega \Big]^2. \tag{2.2}$$

Conversely, any function g that makes the sum of squares small can be thought of as a candidate estimator of f. However, searching for an estimator in this manner constitutes an ill-posed inverse problem in that the solution is unstable and often non-unique. Including a roughness penalty term in (2.2) as a way to regularize the solution is often an effective way to make it a well-posed problem (cf., O'Sullivan (1986) and Tikhonov and Arsenin (1977)). Before we proceed with our approach of regularizing the inverse problem, it is worth pointing out that, even though our aim is to estimate the spectrum on the whole real line, based on partially observed X, it often only makes sense to conduct the estimation on a bounded interval. For instance, if we observe $X(t_i), i = 1, \ldots, N$, where $t_i = k_i\tau$ for some constant $\tau > 0$ and integers k_i, then it is natural to estimate the spectrum only on $[0, \nu]$ with $\nu = \pi/\tau$ to avoid the aliasing effect; see Yaglom (1987, p.187). When $\{t_1, \ldots, t_N\}$ are irregularly spaced but are randomly scattered in $[t_1, t_N]$, a reasonable bounded interval is $[0, \nu]$, where $\nu = \varrho\pi$ and ϱ is the average sampling rate; see Broersen and Bos (2006), Eyer and Bartholdi (1999), and Press et al. (1992). In Eyer and Bartholdi (1999), a case is made that a larger ν should be used in this situation, since the "Nyquist frequency" is usually much larger than the average sampling rate.

To develop our estimator for different ν based on a unified theoretical framework, we first transform the frequency domain from $[0, \nu]$ to $[0, 1]$. Then a possible formulation for the regularized optimization problem described above is to focus on functions g on $[0, 1]$ such that their first $m-1$ derivates $g, g^{(1)}, \ldots, g^{(m-1)}$ are absolutely continuous and $\int_0^1 [g^{(m)}(u)]^2 du < \infty$, for some $m \ge 1$; we look for such g on $[0, 1]$ that solve the minimization problem:

$$\hat{g}_\lambda = \operatorname{argmin}\Big\{ \sum_{1 \le i,j \le N} \Big[X(t_i)X(t_j) - \int_0^1 \cos(\nu(t_i - t_j)u) g(u) du \Big]^2 + \lambda \int_0^1 [g^{(m)}(u)]^2 du \Big\}, \tag{2.3}$$

where λ is an appropriately selected positive constant. Transforming the domain back to $[0, \nu]$, we define the estimator of f as

$$\hat{f}_\lambda(\omega) \equiv \begin{cases} \frac{1}{\nu} \hat{g}_\lambda(\frac{\omega}{\nu}), & \omega \in [0, \nu], \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

The parameter λ controls the smoothness of the solution. A data-driven procedure to select λ is presented in Section 3.2.

In the literature, the space of functions on $[0,1]$ with square-integrable mth derivatives is often denoted by $W_m[0,1]$, or simply W_m, and referred to as a Sobolev space of order m (cf., Wahba (1990)). In this article, we derive the detailed computational and theoretical properties of $\hat{f}_\lambda$ for the case $m=1$. Extending the optimization algorithm to a general m is computationally straightforward, which will be apparent from the derivation of (2.9) below.

From now on, we focus on the case where the optimization in (2.3) is performed with $g \in W_1$. Let $y_i, 1 \le i \le n$, be an arbitrary enumeration of the products of pairs $X(t_j), X(t_k)$, $1 \le j \le k \le N$, where $n = N + N(N-1)/2$, and write $\mathbf{y} \equiv (y_1, \ldots, y_n)^T$; accordingly, let $h_i, 1 \le i \le n$, be the corresponding differences, $|t_j - t_k|$, between the observational time points for the pairs and

$$h_i' \equiv \nu h_i, \ 1 \le i \le n. \tag{2.5}$$

Also, define functions

$$\xi_i(s) \equiv \begin{cases} s - \frac{1}{2}s^2 & h_i = 0, \\ \frac{\sin(h_i')h_i's + \cos(h_i's) - 1}{h_i'^2} & h_i \neq 0, \ 1 \le i \le n, \end{cases} \tag{2.6}$$

and matrices

$$T_{n\times 1} \equiv \Big[\int_0^1 \cos(h_1'u)du, \int_0^1 (\cos(h_2'u)du, \ldots, \int_0^1 \cos(h_n'u)du \Big]^T, \tag{2.7}$$

$$\Sigma_{n\times n} \equiv \Big\{ \int_0^1 \cos(h_i'u)\xi_j(u)du \Big\}_{i,j=1}^n, \quad \text{and} \quad M_{n\times n} \equiv \Sigma + \lambda I, \tag{2.8}$$

where I denotes the identity matrix. Note that T in (2.7) is a matrix, and it should not be confused with T used for the observation interval $[0;T]$; its use should be clear from the context. Then, from the Appendix, it can be seen that

$$\hat{g}_\lambda(u) = d + \sum_{i=1}^n c_i\xi_i(u), \ u \in [0,1], \tag{2.9}$$

where

$$d = (T^T M^{-1} T)^{-1} T^T M^{-1} \mathbf{y}, \tag{2.10}$$

and

$$\mathbf{c} \equiv [c_1, \ldots, c_n]^T = M^{-1}(I - T(T^T M^{-1} T)^{-1} T^T M^{-1})\mathbf{y}. \tag{2.11}$$

In the rest of this paper, we focus on the estimator (2.9). Note that this algorithm in general does not require the data to be observed on a grid, although gridded data do lead to simpler computational formulas. Also, the algorithm can be extended to processes whose index sets are multi-dimensional by working with more general Sobolev spaces.

Theoretically, it is possible for $\hat{f}_\lambda(\omega)$ to exhibit negative values, which we handle by truncation; see below. In all of the examples that we have investigated, this is a minor issue. We demonstrate this point numerically in Section 3.3, by estimating a spectrum that is equal to 0 in a part of the frequency domain. To remove negativity, a practical solution is to threshold $\hat{f}_\lambda(\omega)$ and consider

$$\hat{f}_\lambda^+(\omega) = \max\{\hat{f}_\lambda(\omega), 0\}. \tag{2.12}$$

As seen in Theorems 3 and 5 below, the consistency of $\hat{f}_\lambda(\omega)$ can be established under very general conditions, in which case the difference $\hat{f}_\lambda(\omega)$ and $\hat{f}_\lambda^+(\omega)$ becomes negligible as the sample information increases.

Other approaches to removing negativity include using the constrained optimization algorithm described in Section 9.4 of Wahba (1990). Let $u_1, \ldots, u_\ell$ be a set of points in $[0, 1]$. Adding the constraints $g(u_1) \geq 0, \ldots, g(u_\ell) \geq 0$ in the optimization problem (2.3), the solution can be written as $\check{f}_\lambda(\omega) = (1/\nu)\check{g}_\lambda(\omega/\nu)$, $\omega \in [0, \nu]$, with

$$\check{g}_\lambda(u) = \sum_k b_k\{1 + \min(u_k, u)\} + d + \sum_{i=1}^n c_i\xi_i(u), \ u \in [0, 1],$$

where the constants $\{b_k\}, \{c_i\}, d$ can be obtained by solving a quadratic programming problem. Since the functions in W_1 are smooth, the nonnegativity of $\check{f}_\lambda$ at a suitably selected set of points, $\omega_k = \nu u_k$, virtually guarantees that $\check{f}_\lambda$ is nonnegative everywhere. Compared with $\hat{f}_\lambda$, the price paid for nonnegativity in $\check{f}_\lambda$ is a less-efficient computational algorithm, a lack of deep understanding of how to choose optimally the smoothing parameter, and the absence of a closed-form solution that could be useful for theoretical considerations. The latter two points are amplified in future sections. Another approach to address the lack of complete nonnegativity is to consider the estimation of the logarithm of f. This is common in time series analysis, where the fact that the periodogram values computed at discrete Fourier frequencies are asymptotically uncorrelated and exponentially distributed, makes it natural to conduct approximate likelihood inference on the log spectrum (cf., Pawitan and O'Sullivan (1994)). Such a procedure usually employs numerical optimizations, for which the computations can be quite costly. We do not use these approaches here, although both are worthy of future consideration.

3. Gridded Data

In this section, assume that the observational points $t_1 < \cdots < t_N$ are on the grid $\{k\tau, k = 1, 2, \ldots\}$ for some $\tau > 0$. However, to be consistent with the notation in Section 2, we consider the grid $\mathcal{G} \equiv \{k\pi/\nu, k = 1, 2, \ldots\}$ for some ν, and estimate the spectral density f on $[0, \nu]$. Note that the data are not required to be sequentially observed on $\mathcal{G}$, and hence our results are equally relevant for a spatial process in one dimension. In Section 3.1, Theorem 1 gives a closed-form formula for spectral-density estimation from gridded data. The smoothing parameter selection is considered in Section 3.2, and numerical comparisons with existing time-series procedures for estimating the spectral density are made in Section 3.3. Section 3.4 considers the computation of bounds for the mean squared error. All of the proofs are given in the Appendix.

3.1. A closed-form representation of the estimator

Let $K \equiv (t_N - t_1)\nu/\pi$, $n_k \equiv \sum_{i=1}^{N} \sum_{j=i}^{N} \mathbb{I}(t_j - t_i = k\pi/\nu)$, $k = 0, \ldots, K$, and $n \equiv \sum_{k=0}^{K} n_k$, where $\mathbb{I}$ stands for the indicator function. Define a sequence $\{y_i, i = 1, \ldots, n\}$ as follows. For each $k \geq 0$, define $y_i, \sum_{j=0}^{k-1} n_j + 1 \leq i \leq \sum_{j=0}^{k} n_j$, to be each a product of the form $X(t_u)X(t_v)$ for some $t_u, t_v, 1 \leq u \leq v \leq N$, such that $t_v - t_u = k\pi/\nu$. (For $k = 0$, use the convention that $\sum_{j=0}^{-1} n_j \equiv 0$.) The particular order in which the pairs are indexed within the kth sub-sequence is not important. Thus, for $k = 0$, the n_0 y_i's are equal to the squares of the data; the next n_1 y_i's are products of pairs of data that are observed at distance π/ν apart, and so on. Also recall from Section 2 that $h_i, 1 \leq i \leq n$, is defined as $|t_u - t_v|$ if $y_i = X(t_u)X(t_v)$. Thus, the first n_0 h_i's are all equal to 0, the next n_1 h_i's are all equal to π/ν, and so on. Define

$$S_0 \equiv \sum_{i=1}^{n_0} y_i \quad \text{and} \quad S_k \equiv \sum_{i=\sum_{j=0}^{k-1} n_j + 1}^{\sum_{j=0}^{k} n_j} y_i, \ 1 \leq k \leq K.$$

Let $\hat{f}_\lambda$ be the estimator defined by (2.4).

Theorem 1. *For each $\omega \in [0, \nu]$, we have $\hat{f}_\lambda(\omega) = \sum_{k=0}^{K} b_k(\omega) S_k$, where*

$$b_k(\omega) = \begin{cases} \frac{1}{\nu} \frac{1}{n_0}, & k = 0, \\ \frac{2}{\nu} \frac{\cos(k\pi\omega/\nu)}{n_k + 2(k\pi)^2\lambda}, & k \geq 1. \end{cases}$$

Remark. Suppose now we have time series data where $t_i = i$. The periodogram can be written as $I_T(\omega) = (1/N)S_0 + (2/N) \sum_{k=1}^{K} \cos(k\omega) S_k$. The natural choice of ν in our procedure is $\nu = \pi$. We now compare the formulas for $\hat{f}_\lambda$ and $\pi^{-1} I_T$.

In the summation over $1 \le k \le K$, the coefficient of $\cos(k\omega)S_k$ is $1/[n_k+2(k\pi)^2\lambda]$ for $\hat{f}_\lambda$, where $n_k = N-k$, in comparison to $1/N$ for $\pi^{-1}I_T$. Intuitively, having n_k in the denominator for $\hat{f}_\lambda$ reduces the bias, since that is the number of terms in S_k. While it appears that the other component $2(k\pi)^2\lambda$ in the denominator makes the bias worse as k increases, its real effect is to down-weight the contribution of S_k for large k. This serves to control the variance of $\hat{f}_\lambda$, since the number of terms in S_k tends to decrease as k increases. The manner in which bias and variance are controlled by weights attached to the estimated covariances in our estimator is similar in spirit to the lag-window estimator of the spectral density; see Priestley (1981, pp.432-449), and Brockwell and Davis (1991, pp.351-382). However, the weights in the two procedures are different in form, and, more importantly, the weights in the estimator $\hat{f}_\lambda$ arise naturally from an optimization criterion.

For the gridded data described in this section, consider the modified optimization problem: Find $g \in W_1$ to minimize

$$\sum_{k=1}^{K} n_k\Big(S_k/n_k - \int_0^1 \cos(k\pi u)g(u)du\Big)^2 + \lambda\int_0^1 [g^{(1)}(u)]^2du, \tag{3.1}$$

where $g^{(1)}$ is the first derivative of g. Corollary 2 shows that this leads to precisely $\hat{f}_\lambda$.

Corollary 2. *The estimator of f defined by* (3.1) *and* (2.4) *is equal to $\hat{f}_\lambda$ given by Theorem* 1 *for all $\lambda > 0$.*

3.2. Choosing λ by generalized cross-validation

One approach for selecting λ in $\hat{f}_\lambda$ is generalized cross-validation (GCV); see Chapter 4 of Wahba (1990). The GCV function is ordinarily defined as

$$GCV(\lambda) \equiv \frac{\|(I-H_\lambda)\mathbf{y}\|^2}{[\mathrm{tr}(I-H_\lambda)]^2},\ \lambda > 0, \tag{3.2}$$

where H_λ is the hat matrix (denoted by $A(\lambda)$ in Wahba (1990)) defined by

$$H_\lambda \equiv I - \lambda M^{-1} + \lambda M^{-1}T(T^TM^{-1}T^T)^{-1}T^TM^{-1},$$

and T and M are defined by (2.7) and (2.8) in Section 2. One would then choose λ as the minimizer of $GCV(\lambda)$ given by (3.2). However, as pointed out on p. 65 of Wahba (1990), GCV is likely to give unsatisfactory results when $\{y_i\}$ are highly correlated. This is of course the case, due to the way $\{y_i\}$ are chosen as all possible cross-products. Kohn, Ansley, and Wong (1992) represented a smoothing spline by a state-space model and extended CV, GCV, and GML to

an ARMA setting. Wang (1998) considered smoothing-parameter selection to spatial and spatio-temporal data.

Below we consider two alternatives to minimizing (3.2). Consider the loss $L(\lambda) = (1/n)||\hat{\mathbf{y}} - \mathbb{E}(\mathbf{y})||^2$, where $\hat{\mathbf{y}} = H_\lambda \mathbf{y}$. An unbiased estimator of the risk, $\mathbb{E}[L(\lambda)]$, is

$$V(\lambda) \equiv \frac{1}{n}||(I - H_\lambda)\mathbf{y}||^2 - \frac{1}{n}\mathrm{tr}(\Xi) + \frac{2}{n}\mathrm{tr}(\Xi H_\lambda),$$

where Ξ is an unbiased estimator of the covariance matrix of $\mathbf{y}$. Thus, one can obtain λ as the minimizer of $V(\lambda)$. Another consideration is motivated by the "nil-trace" estimation argument in Li (1985, 1987). Let

$$\alpha \equiv \frac{\mathrm{tr}(\Xi H_\lambda)}{\mathrm{tr}(\Xi) - \mathrm{tr}(\Xi H_\lambda)},$$

and $\tilde{H}_\lambda \equiv -\alpha I + (1+\alpha)H_\lambda$. Note that $\mathrm{tr}(\Xi \tilde{H}_\lambda) = 0$; then

$$\begin{aligned}\tilde{V}(\lambda) &\equiv \frac{1}{n}||(I - \tilde{H}_\lambda)\mathbf{y}||^2 - \frac{1}{n}\mathrm{tr}(\Xi) + \frac{2}{n}\mathrm{tr}(\Xi \tilde{H}_\lambda) \\ &= \frac{\mathrm{tr}(\Xi)}{n^2}\frac{(1/n)||(I - H_\lambda)\mathbf{y}||^2}{[(1/n)\mathrm{tr}(\Xi(I - H_\lambda))]^2} - \frac{1}{n}\mathrm{tr}(\Xi).\end{aligned}$$

One can then choose λ to minimize

$$\widetilde{GCV}(\lambda) \equiv \frac{(1/n)||(I - H_\lambda)\mathbf{y}||^2}{[(1/n)\mathrm{tr}(\Xi(I - H_\lambda))]^2}.$$

In general, having a high-quality unbiased estimator Ξ of the covariance in this problem may be overly ambitious. However, for the situation where data are observed on a grid, as described in Section 3 (or Lemma A3 in the Appendix), it can be seen that H_λ is block-diagonal. As a result, in computing $\mathrm{tr}(\Xi H_\lambda)$ only the corresponding diagonal blocks of Ξ are relevant and, in fact,

$$\mathrm{tr}(\Xi H_\lambda) = \sum_{k=0}^{K} \frac{1}{n_k + 2(k\pi)^2\lambda}\mathrm{tr}(\Xi_k J_{n_k \times n_k}),$$

where Ξ_k is the kth $n_k \times n_k$ diagonal block matrix of Ξ and $J_{n_k \times n_k}$ stands for the matrix of 1's with dimension $n_k \times n_k$. It is clear that Ξ_k is the estimator of the covariance of $y_i, i = \sum_{j=0}^{k-1} n_j + 1, \ldots, \sum_{j=0}^{k} n_j$, which can be obtained through the method of moments.

To see how this works numerically, a simulation study was conducted for the stationary Gaussian process with spectral density,

$$f(\omega) = \begin{cases} 1 & 0 \le \omega < 0.4, \\ \frac{1}{2}\left[1 + \cos\left(\frac{\pi(\omega - 0.4)}{\pi - 0.8}\right)\right] & 0.4 \le \omega \le \pi - 0.4, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

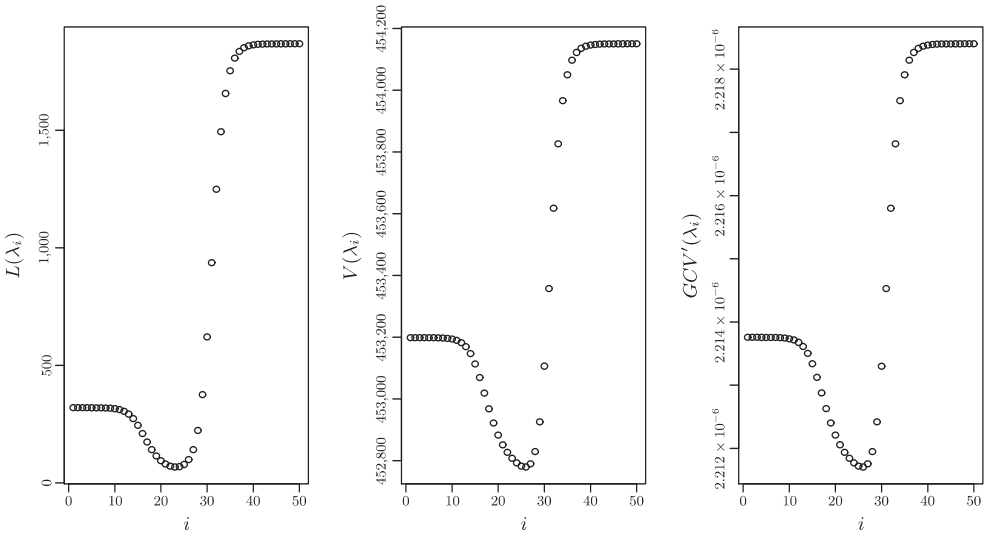

Figure 1. The left plot shows $L(\lambda_i)$ versus i; the middle plot shows $V(\lambda_i)$ versus i; the right plot shows $\widetilde{GCV}(\lambda_i)$ versus i.

Suppose the process is observed at $t = 1, \ldots, 2{,}000$. We illustrate our method for $\lambda_i = 10^{v_i}$, where $v_i = -5 + 14(i-1)/49, 1 \le i \le 50$, based on a single simulation. We computed $\{L(\lambda_i)\}$, $\{V(\lambda_i)\}$, and $\{\widetilde{GCV}(\lambda_i)\}$ and show them plotted against i in Figure 1. It can be seen that the optimal λ determined by these three functions are remarkably close, and $V(\lambda)$ and $\widetilde{GCV}(\lambda)$ are very similar to each other. In the comparisons that follow in the next subsection, we use $\widetilde{GCV}(\lambda)$.

3.3. Comparisons with existing time-series procedures

While our procedure is not restricted to time series, a comparison with leading spectral-density-estimation procedures in the time-series literature is insightful. The procedures chosen here are the following.

Method I: The smoothed periodogram estimator using the Daniell (rectangular) window with the smoothing parameter picked by the cross-validation criterion CVLL, introduced by Beltrão and Bloomfield (1987); see also Hurvich (1985) and Hurvich and Beltrão (1990). Other windows, including the Bartlett, Blackman, Hamming, and Hanning windows, were also tested, but they did not yield better results than the Daniell window.

Method II: The local linear smoother $\hat{m}_{\mathrm{LS}}$ and local maximum likelihood estimator $\hat{m}_{\mathrm{LK}}$, introduced in Fan and Kreutzberger (1998), with smoothing parameters selected by the constant bandwidth selector in Fan and Gijbels (1995).

Table 1. Running Time Comparison (in seconds).

	HHC	BB	FK.LS (FK.LK)	PS
Running time (sec.)	11.34	17.21	69.30	248.35

Method III: The nonparametric approach based on the penalized Whittle likelihood due to Pawitan and O'Sullivan (1994).

We made the comparisons by simulating from a time-series model, and we found overall that our procedure competed well with Methods I, II, and III. We focused on the stationary Gaussian time series with spectral density f given by (3.3), from which we ran 400 simulations of $X(t)$, $t = 1, \ldots, 2{,}000$. The spectral-density estimate $\hat{f}$ was computed for each of five methods, one of which includes our estimator given by Theorem 1. We then computed the sample average of $\{\hat{f}(\omega) - f(\omega)\}^2$ for $\omega \in [0, \pi]$, and the results are displayed in Figure 2. The abbreviations "HHC", "BB", "PS", "FK.LS", and "FK.LK" stand for, respectively, our estimator, the Beltrão and Bloomfield estimator in Beltrão and Bloomfield (1987) (Method I), the Pawitan and O'Sullivan estimator in Pawitan and O'Sullivan (1994) (Method III), and the Fan and Kreutzberger's estimators $\hat{m}_{\mathrm{LS}}$ and $\hat{m}_{\mathrm{LK}}$ in Fan and Kreutzberger (1998) (Method II).

An important point in this comparison is that our procedure is the most computationally efficient, requiring only a small fraction of the computing time required by the Pawitan and O'Sullivan and the Fan and Kreutzberger procedures. A running-time comparison is shown in Table 1 where all four procedures were coded in R (R Development Core Team (2008)); the codes are available upon request. All computations were conducted on a Dual-Core Intel Xeon 1.33 Ghz, Mac Pro machine with 4GB of RAM.

It can be seen from Figure 2 and Table 1 that our estimator performed quite a lot better than the Beltrão and Bloomfield estimator with slightly faster running time, somewhat better than the Fan and Kreutzberger estimators with much faster running time. Finally, our estimator was comparable to the Pawitan and O'Sullivan estimator, but with only a fraction (4.6%) of the running time.

We now revisit the issue that our estimator $\hat{f}_\lambda$ given by (2.9) may take on negative values. Note that the spectrum f used in this example is zero on $[\pi - 0.4, \pi]$. Thus, if the negativity of $\hat{f}_\lambda$ is a prevalent issue in our methodology, we would expect to see a substantial number of negative values for $\hat{f}_\lambda(\omega)$ when ω is close to π. In fact, $\hat{f}_\lambda$ given by (2.9) estimated f exceedingly well close to π; only 18 out of 400 runs produced estimates that had negative values for some ω's. The percentage of negative values in all of the estimates given by (2.9), out of all of the simulations, was .6%, and the minimum of all of the values was

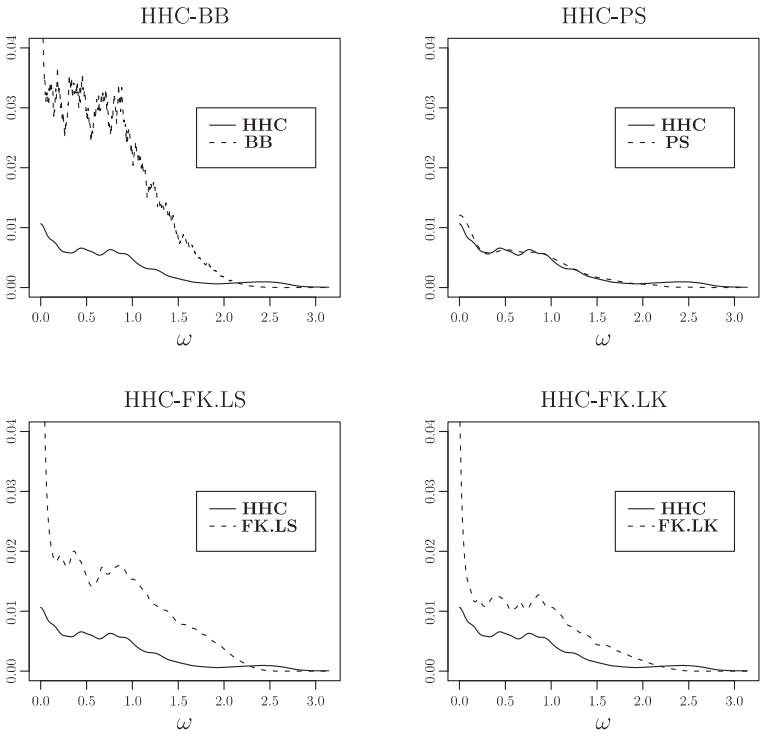

Figure 2. Mean squared error comparisons of our (HHC) estimator with four other estimators. The upper-left plot compares HHC with the Beltrão and Bloomfield (BB) method, the upper-right plot compares HHC with the Pawitan and O'Sullivan (PS) estimator, and the two lower plots compare HHC with the estimators of Fan and Kreutzberger (FK.LS and FK.LK).

-0.050. Thus, the estimator $\hat{f}_\lambda^+$ in (2.12) obtained by thresholding differs little from the original estimator $\hat{f}_\lambda$ in (2.4).

3.4. The mean squared error of the estimator

We assume in this subsection that the data are observed on a grid, and we continue to use the notation developed in Section 3.1. Our goal is to compute bias and variance bounds for $\hat{f}_\lambda(\omega)$. Note that a generic symbol used for a finite positive constant is C, where C may take on different values in different places.

We begin by describing the assumptions. First, the assumption that the stationary process X has a spectral density guarantees that X has the linear-process representation (cf., Yaglom (1987)),

$$X(t) = \int a(t-s)dZ(s),\ t \in \mathbb{R}, \tag{3.4}$$

where $\int a^2(t)dt < \infty$, and Z has stationary uncorrelated increments with mean zero. However, we assume additionally that Z has independent increments; this simplifies the derivations considerably. Let

$$\mathbb{E}[Z(dt)]^2 = dt \quad \text{and} \quad \mathbb{E}[Z(dt)]^4 = \mu_4 dt, \tag{3.5}$$

for some finite μ_4. In the theorems below, these assumptions on X are assumed without further reference.

Also assume that the observational points $t_i, 1 \le i \le N$, are such that $\{n_k\}$ satisfy

$$\inf_{k \le \zeta N} n_k \ge \delta N \tag{3.6}$$

for some $\zeta, \delta \in (0,1)$. The condition (3.6) ensures that there are sufficiently many pairs of data associated with each "small" time lag compared with the sample size. This condition is obviously fulfilled if the data are a time series on $\mathcal{G}$.

The following regularity conditions are also needed.

(C1) Let B be a bounded, symmetric function on $\mathbb{R}$ with $B(t) \downarrow$ for $t > 0$, such that for some $\alpha > 2$ and $C > 0$,

$$B(t) \le Ct^{-\alpha-1}, \text{ for all large } t; \tag{3.7}$$

$$\int |a(u)a(u+t)|du \le B(t), \text{ for all } t; \tag{3.8}$$

and for some $\nu_0 > 0$,

$$\sup_{\nu \ge \nu_0} \frac{1}{\nu} \sum_k |a(\frac{k\pi}{\nu} + u)a(\frac{k\pi}{\nu} + u + t)| \le B(t), \text{ for all } u, t. \tag{3.9}$$

(C2) The covariance function $R(t) \equiv \mathbb{E}[X(0)X(t)]$ is differentiable with $\int |R^{(1)}(t)| dt < \infty$.

Note that $R(t) = \int a(u)a(u+t)dt$ from (3.4) and (3.5). Hence, (3.8) implies that

$$|R(t)| \le B(t), \quad \text{for all } t, \tag{3.10}$$

and (3.7) then implies that X is a short-memory process (cf., Brockwell and Davis (1991, Sec. 13.2)). Observe that the expression on the left of (3.9) approximates $\int |a(u)a(u+t)|du$ if a is continuous and ν_0 is large. Thus, (3.9) is not a strong condition in the presence of (3.8). The condition (C2) requires the covariance function to be sufficiently smooth.

The following result gives bounds for the variance and the absolute bias of $\hat{f}_\lambda$.

Theorem 3. *Assume that* (C1) *and* (C2) *hold. Then there exists a bounded universal constant C such that for all $\nu \geq \nu_0$, $\omega \in [0, \nu]$, N satisfying* (3.6)*, and $\lambda \in [N^{-1}, N]$,*

$$\operatorname{var}(\hat{f}_\lambda(\omega)) \leq C(N\lambda)^{-1/2}, \tag{3.11}$$

$$|\operatorname{bias}(\hat{f}_\lambda(\omega))| \leq C\Big[\nu^{-1} + \frac{\lambda\nu^2}{N} + \Big(\frac{\nu}{N}\Big)^{\alpha}\Big]. \tag{3.12}$$

Corollary 4. *Assume that* (C1) *and* (C2) *hold. Then for*

$$\lambda = \frac{N^{3/5}}{\nu^{8/5}}, \tag{3.13}$$

there exists a bounded universal constant C such that for all $\omega \in [0, \nu]$, N satisfying (3.6)*, and ν satisfying $N \geq \nu \geq \nu_0$,*

$$\operatorname{MSE}(\hat{f}_\lambda(\omega)) \leq C\Big[\Big(\frac{\nu}{N}\Big)^{4/5} + \nu^{-2}\Big]. \tag{3.14}$$

Proof. The proof follows immediately from Theorem 3.

We now discuss the implications for (3.14). Note that $N\pi/\nu$ is roughly the range of the data. As such, (3.14) can be interpreted as

$$\operatorname{MSE}(\hat{f}_\lambda(\omega)) \leq C[(\text{range of data})^{-4/5} + \nu^{-2}].$$

Suppose $X(t)$ is continuously observed for $t \in [0, T]$. Then, taking $N = [T\nu]$ equally spaced observations, where ν satisfies $\nu^{-2} \leq T^{-4/5}$, and for λ given by (3.13), we obtain

$$\operatorname{MSE}(\hat{f}_\lambda(\omega)) \leq CT^{-4/5}.$$

While Theorem 3 and Corollary 4 are proved for the continuous-parameter process (3.4), a quick inspection of the proofs reveals that they also hold for the discrete-parameter process under parallel assumptions. The only major difference is that the absolute bias is bounded by $C(\lambda N^{-1} + N^{-\alpha})$ due to the omission of a Riemann approximation in the derivation. Thus, a MSE rate of $N^{-4/5}$ can be achieved under those assumptions. Note that this coincides with the optimal rate of convergence of the smoothed periodogram estimator; see the discussions on pp. 567-568 of Priestley (1981) and in Section 4.7 of Grenander and Rosenblatt (1984).

Theorem 3 can be generalized in a number of ways, including relaxing (3.6) and conditions (C1) and (C2), or going beyond the linear process (3.4). These extensions, while useful, make the proofs longer and more technical. In Section 4, we consider the more pertinent extension of allowing $\{t_i\}$ to be nongridded.

4. Non-gridded Data

We have demonstrated that our methodology competes well with periodogram-based approaches in the context of time series data. However, one enormous advantage of our approach is the flexibility it offers in terms of the wide range of spectral estimation problems that we can readily address. In this section, the same approach is applied when the observational points $t_1, \ldots, t_N$ are not on a grid. Spectral-density estimation in the case of non-gridded data has received relatively little attention. Note that non-gridded data can always be presented as approximately gridded data with missing values. As such, any procedure for gridded data that allows missing values can potentially be used for this situation; see, for instance, Broersen and Bos (2006), Bos, de Waele, and Broersen (2002), Broersen (2008), Lomb (1976), Scargle (1992), and Stoica and Sandgren (2006). Our procedure intrinsically does not distinguish between gridded and non-gridded data; indeed the general formula (2.9) can be readily used to handle both data scenarios. However, in the cases of non-gridded data, applying (2.9) lacks computational efficiency since matrices of size $O(N^2) \times O(N^2)$ are involved. We present an approach that approximates non-gridded data by gridded data below.

The closed-form formula in Theorem 1 can be used by projecting the irregularly observed points onto the nearest grid $\mathcal{G} \equiv \{k\pi/\nu, k = 1, 2, \ldots\}$. Should $\{t_i\}$ be randomly scattered in $[t_1, t_N]$, in accordance with the discussion on choice of ν in Section 2, we recommend choosing ν to be three to four times $\varrho\pi$, where ϱ is the average sampling rate. Numerical studies later in this section show that the estimation result is not sensitive to this choice. For each $k = 0, 1, \ldots$, let $L_k = \Big\{(t_i, t_j) : t_i \in k_i\pi/\nu \pm \pi/(2\nu),\ t_j \in k_j\pi/\nu \pm \pi/(2\nu),\ |k_i - k_j| = k\Big\}$ and $n'_k = |L_k|$, where, for convenience, $k_i\pi/\nu \pm \pi/(2\nu)$ denotes the interval $(k_i\pi/\nu - \pi/(2\nu), k_i\pi/\nu + \pi/(2\nu)]$. Then, by Theorem 1, an estimator of f is

$$\tilde{f}_\lambda(\omega) = \frac{1}{\nu}\frac{1}{n'_0}S'_0 + \frac{2}{\nu}\sum_{k=1}^{K}\frac{\cos(k\pi\omega/\nu)}{n'_k + 2(k\pi)^2\lambda}S'_k, \tag{4.1}$$

where $S'_k = \sum_{(t_i,t_j)\in L_k} X(t_i)X(t_j)$.

To evaluate the bias and variance bounds for the spectrum estimator, we need first to strengthen the conditions (C1) and (C2), as follows.

(C1′) Let B be a bounded, symmetric function on $\mathbb{R}$ with $B(t) \downarrow$ for $t > 0$, such that for some $\alpha > 2$ and $C > 0$,

$$B(t) \le Ct^{-\alpha-1}, \text{ for all } t \text{ large.}$$

Let $\beta(t) \equiv \sup_{|\delta| \le \pi/\nu_0} |a(t+\delta)|$, for some $\nu_0 > 0$. Then B satisfies

$$\int \beta(u)\beta(u+t)du \le B(t), \text{ for all } t, \tag{4.2}$$

$$\sup_{\nu \ge \nu_0} \nu^{-1} \sum_k \beta(\frac{k\pi}{\nu} + u)\beta(\frac{k\pi}{\nu} + u + t) \le B(t), \text{ for all } u, t.$$

(C2′) The covariance function $R(t)$ is differentiable. If $Q(t) \equiv \sup_{|\delta| \le \pi/\nu_0} |R^{(1)}(t+\delta)|$, then $\int Q(t)dt < \infty$.

Let $\rho_\nu \equiv \sup_k \#\{t_i : t_i \in k\pi/\nu \pm \pi/(2\nu)\}$, the maximum number of $\{t_i\}$ projected to any grid point. The following result gives bounds for the variance and the absolute bias of $\tilde{f}_\lambda$.

Theorem 5. *Assume that* (C1′) *and* (C2′) *hold. Then there exists a bounded universal constant C such that for all $\nu \ge \nu_0$, $\omega \in [0, \nu]$, N satisfying* (3.6)*, and $\lambda \in [N^{-1}, N]$,*

$$\operatorname{var}(\tilde{f}_\lambda(\omega)) \le C\rho_\nu(N\lambda)^{-1/2},$$
$$|\operatorname{bias}(\tilde{f}_\lambda(\omega))| \le C\Big[\nu^{-1} + \frac{\lambda\nu^2}{N} + \Big(\frac{\nu}{N}\Big)^\alpha\Big].$$

Observe that ρ_ν is bounded if the $\{t_i\}$ are randomly scattered and ν is larger than the average sampling rate. In that case, the variance and absolute bias bounds are identical to those for the gridded case in (3.11) and (3.12) and as a consequence, the same MSE rate described in Corollary 4 holds. To the best of our knowledge, precise MSE-rate computations for other procedures for non-gridded data are not available.

To demonstrate how $\tilde{f}_\lambda$ works, we consider estimation of the spectral density of a continuous-parameter Gaussian process whose values are observed at random time points (or locations). Specifically, we assumed the spectral density to be $f(\omega) = 10(1+\omega^2)^{-1}$, $\omega \in \mathbb{R}$. In our simulations, the process was observed at 2,000 points that were i.i.d. uniformly distributed on [0, 1,000]. We then carried out the analysis using the estimator (4.1), with smoothing parameter λ selected by $\widetilde{GCV}(\lambda)$. Note that in this case, we could have multiple observations, or no observation at all, at any grid point. Recall our recommendation that ν be 3 to 4 times $\rho\pi$, where ρ is the average sampling rate. In this example, ρ is about 2, and hence to assess the choice of ν we considered $\tilde{f}_\lambda$ for $\nu = 2\pi$, 4π, 6π, and 10π, based on one simulation. The results are plotted in Figure 3; for the range of ν considered, the performance of the estimator was quite stable.

We next compared our approach with other approaches for non-gridded data in the literature. We focused on the comparison with a procedure due to P.

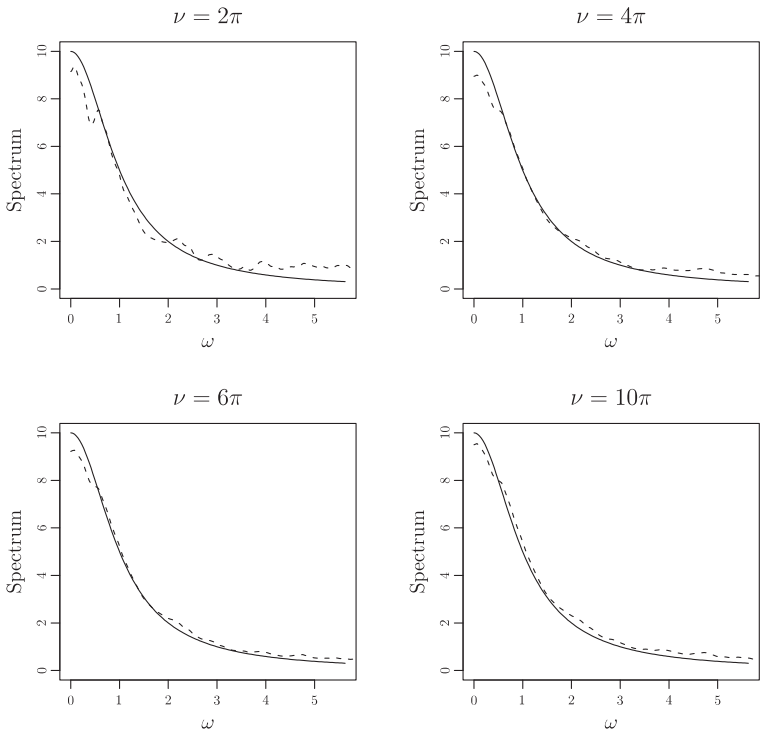

Figure 3. Plots of true spectral density (solid line) and estimated spectral density $\tilde{f}_\lambda$ (dashed line) using different choices of ν.

Broersen, introduced in Broersen and Bos (2006); we did not compare with the Lomb-Scargle periodogram, as it is mainly intended for detecting frequencies. To implement Broersen's method, we used his MATLAB package available at MATLAB Central. The results for analyzing one simulation are presented in the left plot of Figure 4. From 100 simulations, the average squared prediction error was computed for each estimator as a function of ω. The comparison is shown in the right plot of Figure 4. Based on these simulation results, our method did better overall and, especially so for smaller values of ω.

5. Discussion and Conclusions

We have limited our attention to stationary processes on $\mathbb{R}$. The formulation of our methodology can, in principle, be adapted for spectral-density estimation of spatial processes that are stationary, or are intrinsic random functions. These extensions will be investigated in future work. We restricted our estimator to be in the Sobolev space W_1 to keep technical derivations and the requirement on the smoothness of the spectral density to a minimum. The theoretical properties

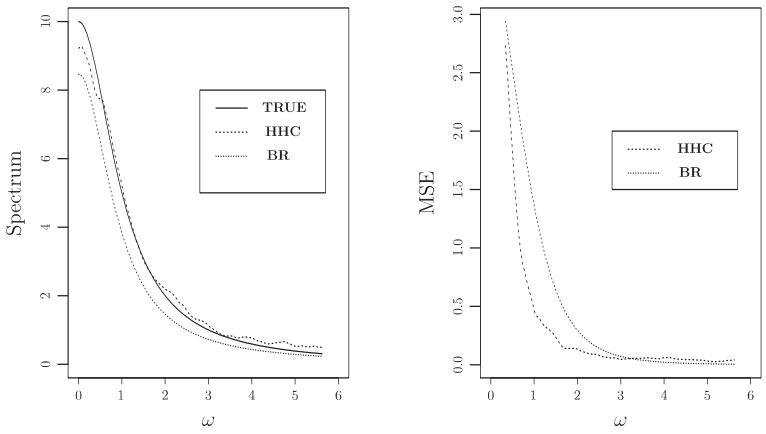

Figure 4. Plots of true spectral density (solid line), our (HHC) estimate $\tilde{f}_\lambda$ (dashed line), and Broersen's (BR) estimate (dotted line). The right plot compares the mean squared error for HHC and BR based on 100 simulations.

of the estimator when the space is taken to be a Sobolev space of higher order would be an interesting but complicated extension.

In conclusion, we have described a new methodology for estimating the spectral density function of a stationary process based on a regularized inverse problem. The methodology

1. does not require the data to be observed on a regular grid, and holds strong promise for being adaptable to more general settings such as for intrinsic random functions (cf., Matheron (1973));
2. is computationally and statistically efficient;
3. does not make use of the periodogram in the time-series setting (where it performs at least as well as, and sometimes better than bench-mark periodogram-based procedures).

Acknowledgements

We are grateful to an associate editor and a referee for their careful reading of the paper and helpful remarks. This work was supported by NSF Grants DMS-0808993 and DMS-0806098, ONR Grant N00014-08-01-0464, and UT-Battelle LLC contract 4000042084.

Appendix

Proof of (2.9). This proof is an application of Theorem 1.3.1 of Wahba (1990), to which readers are referred for details. We briefly review the background.

Define an inner product on W_1 as

$$\langle g_1, g_2\rangle_{W_1} \equiv g_1(0)g_2(0) + \int_0^1 g_1'(u)g_2'(u)du,\ g_1, g_2 \in W_1.$$

Let $\mathcal{H}_0$ be the space of constant functions on $[0,1]$, and $\mathcal{H}_1$ be the space of absolutely continuous, measurable functions q on $[0,1]$ with $q(0)=0$ and $\int_0^1 [q^{(1)}(u)]^2 du < \infty$; also make $\mathcal{H}_0$ and $\mathcal{H}_1$ Hilbert spaces by defining the inner products $\langle p_1, p_2\rangle_{\mathcal{H}_0} = p_1(0)p_2(0), p_1, p_2 \in \mathcal{H}_0$ and $\langle q_1, q_2\rangle_{\mathcal{H}_1} = \int_0^1 q_1^{(1)}(u)q_2^{(1)}(u)du, q_1, q_2 \in \mathcal{H}_1$. It is easy to see that each function g in W_1 can be uniquely written as $g = p + q$, for some $p \in \mathcal{H}_0$ and $q \in \mathcal{H}_1$. This decomposition is often described by the notation $W_1 = \mathcal{H}_0 \oplus \mathcal{H}_1$. Note that W_1 is a reproducing kernel Hilbert space, since

$$\langle g, \mathcal{R}(\cdot, t)\rangle_{W_1} = g(t),\ \text{for each } t \in [0,1], \tag{A.1}$$

where the bivariate function $\mathcal{R}(s,t) \equiv 1 + \min(s,t),\ s,t \in [0,1]$, is called the reproducing kernel and (A.1) is referred to as the reproducing property. Let $\mathcal{R}_s = \mathcal{R}(s,\cdot)$. Define the bounded linear functional, $L_i g \equiv \int_0^1 \cos(h_i' u)g(u)du, g \in W_1$, where h_i' is defined by (2.5); let η_i be the representer of L_i, so that $L_i g = \langle \eta_i, g\rangle_{W_1}, g \in W_1$. By the reproducing property,

$$\begin{aligned}
\eta_i(s) = \langle \eta_i, \mathcal{R}_s\rangle_{W_1} = L_i\mathcal{R}_s &= \int_0^1 \cos(h_i' u)(1 + \min(s,u))du \\
&= \int_0^1 \cos(h_i' u)du + \int_0^s \cos(h_i' u)udu + \int_s^1 \cos(h_i' u)sdu \\
&= \frac{\sin(h_i')}{h_i'} + \frac{(\sin h_i')h_i' s + \cos(h_i' s) - 1}{h_i'^2}.
\end{aligned}$$

It is clear that $\xi_i(s)$ defined by (2.6) is the projection of η_i on $\mathcal{H}_1$. Then (2.9) follows from a direct application of Theorem 1.3.1 of Wahba (1990).

Proof of Theorem 1. To make the proof more readable, we include the technical details in a few lemmas below. In the following, let I_k be the identity matrix of dimension $k \times k$, and let $J_{k_1 \times k_2}$ be the matrix of 1's with dimension $k_1 \times k_2$. Define the matrices:

$$B_k \equiv \Big[\frac{1}{2[(k-1)\pi]^2} J_{n_{k-1}\times n_{k-1}} + \lambda I_{n_{k-1}}\Big]^{-1},\ 2 \le k \le K+1, \tag{A.2}$$

$$A_1 \equiv I_{n_0},\ A_k \equiv \frac{1}{[(k-1)\pi]^2} B_k J_{n_{k-1}\times n_0},\ 2 \le k \le K+1, \tag{A.3}$$

$$A_0 \equiv \Big[\frac{1}{3} J_{n_0\times n_0} + \lambda I_{n_0}\Big] - \sum_{k=2}^{K+1} \frac{1}{[(k-1)\pi]^2} J_{n_0\times n_{k-1}} A_k, \tag{A.4}$$

and write

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{K+1} \end{bmatrix}.$$

First note that by the identity

$$\Big(\frac{1}{c}J_{k\times k} + I_k\Big)^{-1} = I_k - \frac{1}{k+c}J_{k\times k},\ c \neq 0, \tag{A.5}$$

we obtain

$$B_k = \lambda^{-1}\Big(I_{n_{k-1}} - \frac{1}{n_{k-1} + 2[(k-1)\pi]^2\lambda}J_{n_{k-1}\times n_{k-1}}\Big), \tag{A.6}$$

which implies that

$$A_k = \frac{1}{[(k-1)\pi]^2}B_k J_{n_{k-1}\times n_0} = \frac{2}{n_{k-1} + 2[(k-1)\pi]^2\lambda}J_{n_{k-1}\times n_0}\ . \tag{A.7}$$

Lemma A.1.

$$A_0^{-1} = \lambda^{-1}I_{n_0} - \lambda^{-1}\frac{1}{n_0 + \lambda/b_0}J_{n_0\times n_0}, \tag{A.8}$$

where

$$b_0 \equiv \frac{1}{3} - 2\sum_{k=2}^{K+1}\frac{1}{[(k-1)\pi]^2}\frac{n_{k-1}}{n_{k-1} + 2[(k-1)\pi]^2\lambda}\ . \tag{A.9}$$

Proof. By (A.4) and (A.7),

$$\begin{aligned} A_0 &= \Big[\frac{1}{3}J_{n_0\times n_0} + \lambda I_{n_0}\Big] - \sum_{k=2}^{K+1}\frac{1}{[(k-1)\pi]^2}J_{n_0\times n_{k-1}}A_k \\ &= \Big[\frac{1}{3}J_{n_0\times n_0} + \lambda I_{n_0}\Big] - 2\sum_{k=2}^{K+1}\frac{1}{[(k-1)\pi]^2}\frac{n_{k-1}}{n_{k-1} + 2[(k-1)\pi]^2\lambda}J_{n_0\times n_0} \\ &= \lambda I_{n_0} + b_0 J_{n_0\times n_0}\ . \end{aligned}$$

The result then follows from (A.5).

Lemma A.2. $M^{-1} = \mathrm{diag}\{0, B_2, \ldots, B_{K+1}\} + AA_0^{-1}A^T$.

Proof. Recall the definition of Σ and M in (2.8). We first compute Σ. It follows from (2.6) that

$$\xi_i^{(1)}(s) = \begin{cases} 1-s & h_i' = 0, \\ \frac{\sin(h_i') - \sin(h_i' s)}{h_i'} & h_i' \neq 0. \end{cases}$$

Straightforward calculations show that

$$\begin{aligned}
\langle \xi_i, \xi_j \rangle_{W_1} &= \int_0^1 (1-s)^2 ds = \frac{1}{3}, \ h_i' = h_j' = 0, \\
\langle \xi_i, \xi_j \rangle_{W_1} &= -\int_0^1 (1-s) \frac{\sin(k\pi s)}{k\pi} ds = -\frac{1}{(k\pi)^2}, \ h_i' = 0, h_j' = k\pi, k \neq 0, \\
\langle \xi_i, \xi_j \rangle_{W_1} &= \int_0^1 \Big(\frac{\sin(k\pi s)}{k\pi} \Big)^2 ds = \frac{1}{2(k\pi)^2}, \ h_i' = h_j' = k\pi, k \neq 0, \\
\langle \xi_i, \xi_j \rangle_{W_1} &= \int_0^1 \Big(\frac{\sin(k_1\pi s)}{k_1\pi} \Big) \Big(\frac{\sin(k_2\pi s)}{k_2\pi} \Big) ds = 0, \\
&\qquad\qquad h_i' = k_1\pi, h_j' = k_2\pi, 0 \neq k_1 \neq k_2 \neq 0.
\end{aligned}$$

It follows that Σ is symmetric and has a block structure: $\Sigma = \{\Sigma_{i,j}\}_{i,j=0}^K$, where

$$\begin{aligned}
\Sigma_{0,0} &= \frac{1}{3} J_{n_0 \times n_0}, \\
\Sigma_{0,k} &= -\frac{1}{k\pi^2} J_{n_0 \times n_k}, \ 1 \leq k \leq K, \\
\Sigma_{k,k} &= \frac{1}{2\pi^2} J_{n_k \times n_k}, \quad 1 \leq k \leq K, \\
\Sigma_{i,j} &= 0, \text{ otherwise.}
\end{aligned}$$

We now compute $M^{-1} \equiv \{M^{ij}\}$, where the M^{ij} are the block matrices corresponding to the blocks in Σ. To solve for $M^{k1}, 1 \leq k \leq K+1$, the first column (or row) of M^{-1}, we note that

$$\begin{bmatrix} I_{n_0} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = (\Sigma + \lambda I) \begin{bmatrix} M^{11} \\ M^{21} \\ \vdots \\ M^{(K+1),1} \end{bmatrix},$$

which gives

$$\begin{aligned}
I_{n_0} &= \Big(\frac{1}{3} J_{n_0 \times n_0} + \lambda I_{n_0} \Big) M^{11} - \frac{1}{\pi^2} J_{n_0 \times n_1} M^{21} - \ldots - \frac{1}{(K\pi)^2} J_{n_0 \times n_K} M^{(K+1),1}, \\
0 &= -\frac{1}{\pi^2} J_{n_1 \times n_0} M^{11} + \Big(\frac{1}{2\pi^2} J_{n_1 \times n_1} + \lambda I_{n_1} \Big) M^{21},
\end{aligned}$$

$$\vdots$$
$$0 = -\frac{1}{(K\pi)^2} J_{n_K \times n_0} M^{11} + \Big(\frac{1}{2(K\pi)^2} J_{n_K \times n_K} + \lambda I_{n_K}\Big) M^{(K+1),1}.$$

It is clear that $M^{k1} = A_k M^{11}$, $2 \le k \le K+1$, and $M^{11} = A_0^{-1}$. Proceeding in this manner, in general we have

$$\begin{aligned} M^{kj} &= A_k A_0^{-1} A_j^T, \quad k \ne j, \\ M^{11} &= A_0^{-1}, \\ M^{kk} &= B_k + A_k A_0^{-1} A_k^T, \quad 2 \le k \le K+1. \end{aligned}$$

By (2.7) and the fact that $\{h_i\}$ is on a regular grid, we have

$$T = \begin{bmatrix} \mathbf{1}_{n_0} \\ \mathbf{0}_{n-n_0} \end{bmatrix}. \tag{A.10}$$

Let $H_\lambda = I - \lambda M^{-1} + \lambda M^{-1} T (T^T M^{-1} T^T)^{-1} T^T M^{-1}$, so that from (2.11),

$$\mathbf{c} = \lambda^{-1}(I - H_\lambda)\mathbf{y}. \tag{A.11}$$

Lemma A.3.
1. *With b_0 is given by* (A.9),

$$a_0 \equiv (T^T M^{-1} T)^{-1} = \lambda \frac{n_0 + \lambda/b_0}{n_0 \lambda / b_0}.$$

2. $$\begin{aligned} H_\lambda &= \operatorname{diag}\{n_0^{-1} J_{n_0 \times n_0}, I_{n_1} - \lambda B_2, \ldots, I_{n_K} - \lambda B_{K+1}\} \\ &= \operatorname{diag}\Big\{\frac{1}{n_0} J_{n_0 \times n_0}, \frac{1}{n_1 + 2\pi^2 \lambda} J_{n_1 \times n_1}, \ldots, \frac{1}{n_K + 2(K\pi)^2 \lambda} J_{n_K \times n_K}\Big\}. \end{aligned}$$

Proof. Since $A_1 = I_{n_0}$, it follows from (A.10) and Lemma A.2 that $a_0 = (\mathbf{1}_{n_0}^T A_0^{-1} \mathbf{1}_{n_0})^{-1}$. Hence, (i) follows simply from Lemma A.1 and (A.5). Similarly, we have

$$M^{-1} T (T^T M^{-1} T)^{-1} T^T M^{-1} = a_0 A A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A^T.$$

By this and Lemma A.2,

$$\begin{aligned} H_\lambda &= I - \lambda \operatorname{diag}\{0, B_2, \ldots, B_{K+1}\} - \lambda A A_0^{-1} A^T + \lambda a_0 A A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A^T \\ &= \operatorname{diag}\{I_{n_0}, I_{n_1} - \lambda B_2, \ldots, I_{n_K} - \lambda B_{K+1}\} + \lambda \tilde{A}, \end{aligned} \tag{A.12}$$

where

$$\begin{aligned} \tilde{A} &= -A A_0^{-1} A^T + a_0 A A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A^T \\ &= \{-A_i A_0^{-1} A_j^T + a_0 A_i A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T\}_{i,j=1,\ldots,K+1}. \end{aligned}$$

If at least one of i, j is greater than or equal to 2, say $i \geq 2$, then by part (i) (already proved) and (A.3),

$$\begin{aligned} a_0 A_i A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T &= \frac{a_0}{[(i-1)\pi]^2} B_i J_{n_{i-1}\times n_0} A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T \\ &= \frac{a_0}{[(i-1)\pi]^2} B_i \mathbf{1}_{n_{i-1}} \mathbf{1}_{n_0}^T A_0^{-1} \mathbf{1}_{n_0} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T \\ &= \frac{a_0}{[(i-1)\pi]^2} B_i \mathbf{1}_{n_{i-1}} a_0^{-1} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T \\ &= \frac{1}{[(i-1)\pi]^2} B_i \mathbf{1}_{n_{i-1}} \mathbf{1}_{n_0}^T A_0^{-1} A_j^T = A_i A_0^{-1} A_j^T. \end{aligned}$$

This shows that the blocks in the matrix $\tilde{A}$ are all zero except for the first block $\tilde{A}_{11}$ with size $n_0 \times n_0$. It is then easy to verify, using (A.8) and part (i) of this lemma, that $\tilde{A}_{11} = -\lambda^{-1} I_{n_0} + \lambda^{-1} n_0^{-1} J_{n_0\times n_0}$. This shows that $\lambda\tilde{A} = \operatorname{diag}\{-I_{n_0} + n_0^{-1} J_{n_0\times n_0}, 0, \cdots, 0\}$. Thus, (ii) follows from (A.12) and (A.6).

By (i) of Lemma A.3 and Lemma A.1,

$$\begin{aligned} a_0 \mathbf{1}_{n_0}^T A_0^{-1} &= \lambda \frac{n_0 + \lambda/b_0}{n_0\lambda/b_0} \mathbf{1}_{n_0}^T \Big(\lambda^{-1} I_{n_0} - \lambda^{-1} \frac{1}{n_0 + \lambda/b_0} J_{n_0\times n_0}\Big) \\ &= \frac{n_0 + \lambda/b_0}{n_0\lambda/b_0} \Big(1 - \frac{n_0}{n_0 + \lambda/b_0}\Big) \mathbf{1}_{n_0}^T = n_0^{-1} \mathbf{1}_{n_0}^T, \end{aligned}$$

and by (A.7),

$$a_0 \mathbf{1}_{n_0}^T A_0^{-1} A_{k+1}^T = n_0^{-1} \mathbf{1}_{n_0}^T J_{n_0\times n_k} \frac{2}{n_k + 2(k\pi)^2\lambda} = \frac{2}{n_k + 2(k\pi)^2\lambda} \mathbf{1}_{n_k}^T.$$

It now follows from (2.10) that

$$d = (T^T M^{-1} T)^{-1} T^T M^{-1} \mathbf{y} = a_0 \mathbf{1}_{n_0}^T A_0^{-1} A^T \mathbf{y} = \frac{1}{n_0} S_0 + 2 \sum_{k=1}^{K} \frac{1}{n_k + 2(k\pi)^2\lambda} S_k,$$

where $\mathbf{y}_k$ is the $n_k \times 1$ vector of y's that correspond to pairs of $t_i < t_j$ with $t_j - t_i = k\pi/\nu$. By (A.11), (ii) of Lemma A.3, and (2.6),

$$\sum_i c_i \xi_i(u) = \lambda^{-1} \mathbf{1}_{n_0}^T (I_{n_0} - n_0^{-1} J_{n_0\times n_0}) \mathbf{y}_0 (u - \frac{u^2}{2}) + \sum_{k=1}^{K} \mathbf{1}_{n_k}^T B_{k+1} \mathbf{y}_k \frac{\cos(k\pi u) - 1}{(k\pi)^2}.$$

The first term is equal to 0, whereas, by (A.6), the second term is equal to

$$\sum_{k=1}^{K} \mathbf{1}_{n_k}^T \Big(\lambda^{-1} I_{n_k} - \lambda^{-1} \frac{1}{n_k + 2(k\pi)^2\lambda} J_{n_k\times n_k}\Big) \mathbf{y}_k \frac{\cos(k\pi u) - 1}{(k\pi)^2}$$

$$
\begin{aligned}
&= \sum_{k=1}^{K} \Big(\lambda^{-1} S_k - \lambda^{-1} \frac{n_k}{n_k + 2(k\pi)^2 \lambda} S_k \Big) \frac{\cos(k\pi u) - 1}{(k\pi)^2} \\
&= 2 \sum_{k=1}^{K} \frac{(k\pi)^2}{n_k + 2(k\pi)^2 \lambda} S_k \frac{\cos(k\pi u) - 1}{(k\pi)^2} \\
&= 2 \sum_{k=1}^{K} \frac{1}{n_k + 2(k\pi)^2 \lambda} S_k (\cos(k\pi u) - 1).
\end{aligned}
$$

Combining d and $\sum_i c_i \xi_i(u)$ and making the transformation (2.4), we obtain the formula for $\hat{f}_\lambda(\omega)$ in Theorem 1. This concludes the proof of Theorem 1.

Proof of Corollary 2. The proof follows from a simple adaptation of the proof of Theorem 1. Redefine the first n_0 y's to be all equal to S_0/n_0, the next n_1 y's to be all equal to S_1/n_1, and so on.

Proof of Theorem 3. Recall the notation that C is a generic symbol for a finite positive constant (whose value may be different in different places). We first prove the bound for the variance. Write

$$
\operatorname{var}(\hat{f}_\lambda(\omega)) = \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} b_{k_1} b_{k_2} \operatorname{cov}(S_{k_1}, S_{k_2}),
$$

where we suppress the argument ω in b_k. For $1 \le k \le K$, define $w_k(i,j) = \mathbb{I}(t_j - t_i = k\pi/\nu)$, $1 \le i \le j \le N$. Let $\mathcal{I}_k = \{1 \le i \le N : \sum_{j=i}^{N} w_k(i,j) > 0\}$, so that $S_k = \sum_{i \in \mathcal{I}_k} \sum_{j=i}^{N} w_k(i,j) X(t_i) X(t_j)$. Then

$$
\begin{aligned}
&\operatorname{cov}(S_{k_1}, S_{k_2}) \\
&\quad = \sum_{i_1 \in \mathcal{I}_{k_1}} \sum_{i_2 \in \mathcal{I}_{k_2}} \sum_{j_1=i_1}^{N} \sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1) w_{k_2}(i_2,j_2) \operatorname{cov}(X(t_{i_1})X(t_{j_1}), X(t_{i_2})X(t_{j_2})) \\
&\quad = \sum_{i_1 \in \mathcal{I}_{k_1}} \sum_{i_2 \in \mathcal{I}_{k_2}} \sum_{j_1=i_1}^{N} \sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1) w_{k_2}(i_2,j_2) \\
&\qquad \times \Big\{ \mathbb{E}[X(t_{i_1})X(t_{j_1})X(t_{i_2})X(t_{j_2})] - \mathbb{E}[X(t_{i_1})X(t_{j_1})] \mathbb{E}[X(t_{i_2})X(t_{j_2})] \Big\} \\
&\quad = A_1 - A_2,
\end{aligned}
$$

where

$$
A_1 = \sum_{i_1 \in \mathcal{I}_{k_1}} \sum_{i_2 \in \mathcal{I}_{k_2}} \sum_{j_1=i_1}^{N} \sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1) w_{k_2}(i_2,j_2)
$$

$$\times \iiint\!\!\int a(t_{i_1}-u_1)a(t_{j_1}-u_2)a(t_{i_2}-u_3)a(t_{j_2}-u_4)$$
$$\times \mathbb{E}[Z(du_1)Z(du_2)Z(du_3)Z(du_4)],$$
$$A_2 = \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\sum_{j_1=i_1}^{N}\sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1)w_{k_2}(i_2,j_2)R(t_{i_1}-t_{j_1})R(t_{i_2}-t_{j_2}).$$

Write $g(u;i,k)=\sum_{j=i}^{N} w_k(i,j)a(t_j-u)$. Clearly,

$$|g(u;i,k)| \le |a(t_i+\frac{k\pi}{\nu}-u)|. \tag{A.13}$$

Recall that Z has independent increments. Hence, we decompose A_1 into four terms: $A_1 = A_{11}+A_{12}+A_{13}+A_{14}$, where

$$A_{11} = \mu_4 \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\int_u a(t_{i_1}-u)g(u;i_1,k_1)a(t_{i_2}-u)g(u;i_2,k_2)du,$$
$$A_{12} = \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\int_{u_1}\int_{u_2} a(t_{i_1}-u_1)g(u_1;i_1,k_1)a(t_{i_2}-u_2)g(u_2;i_2,k_2)du_1du_2,$$
$$A_{13} = \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\int_{u_1}\int_{u_2} a(t_{i_1}-u_1)g(u_2;i_1,k_1)a(t_{i_2}-u_1)g(u_2;i_2,k_2)du_1du_2,$$
$$A_{14} = \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\int_{u_1}\int_{u_2} a(t_{i_1}-u_1)g(u_2;i_1,k_1)a(t_{i_2}-u_2)g(u_1;i_2,k_2)du_1du_2.$$

Clearly, $A_{12}=A_2$. By the triangle inequality, (A.13), and condition (C1),

$$\begin{aligned}|A_{11}| &\le \mu_4 \sum_{i_1\in\mathcal{I}_{k_1}}\int_u |a(t_{i_1}-u)g(u;i_1,k_1)|\sum_{i_2\in\mathcal{I}_{k_2}}|a(t_{i_2}-u)g(u;i_2,k_2)|du\\ &\le C\sum_{i_1\in\mathcal{I}_{k_1}}\int_u |a(t_{i_1}-u)a(t_{i_1}+\frac{k_1\pi}{\nu}-u)|\sum_{i_2\in\mathcal{I}_{k_2}}|a(t_{i_2}-u)a(t_{i_2}+\frac{k_2\pi}{\nu}-u)|du\\ &\le C\nu\sum_{i_1\in\mathcal{I}_{k_1}} B(\frac{k_1\pi}{\nu})B(\frac{k_2\pi}{\nu}) = C\nu n_{k_1}B(\frac{k_1\pi}{\nu})B(\frac{k_2\pi}{\nu}).\end{aligned}$$

Similarly,

$$\begin{aligned}|A_{13}| \le \sum_{i_1\in\mathcal{I}_{k_1}}\sum_{i_2\in\mathcal{I}_{k_2}}\int_{u_1}\int_{u_2} &|a(t_{i_1}-u_1)a(t_{i_1}+\frac{k_1\pi}{\nu}-u_2)|\\ &\times|a(t_{i_2}-u_1)a(t_{i_2}+\frac{k_2\pi}{\nu}-u_2)|du_1du_2\end{aligned}$$

$$\leq C \sum_{i\in\mathcal{I}_{k_1}} \sum_{i_2\in\mathcal{I}_{k_2}} B(t_{i_1}-t_{i_2})B(t_{i_1}-t_{i_2}+(k_1-k_2)\frac{\pi}{\nu})$$
$$\leq Cn_{k_1} \sum_i B(t_i)B(t_i+(k_1-k_2)\frac{\pi}{\nu}),$$

$$|A_{14}| \leq Cn_{k_1} \sum_i B(t_i+\frac{k_1\pi}{\nu})B(t_i-\frac{k_2\pi}{\nu}),$$

Thus, $\mathrm{var}(\hat{f}_\lambda(\omega)) \leq T_1+T_2+T_3$, where

$$T_1 = C\nu \sum_{k_1=0}^{K} n_{k_1}|b_{k_1}|B(\frac{k_1\pi}{\nu}) \sum_{k_2=0}^{K} |b_{k_2}|B(\frac{k_2\pi}{\nu}),$$
$$T_2 = C \sum_{k_1=0}^{K} n_{k_1}|b_{k_1}| \sum_i B(t_i) \sum_{k_2=0}^{K} |b_{k_2}|B(t_i+(k_1-k_2)\frac{\pi}{\nu}),$$
$$T_3 = C \sum_{k_1=0}^{K} n_{k_1}|b_{k_1}| \sum_i B(t_i+\frac{k_1\pi}{\nu}) \sum_{k_2=0}^{K} |b_{k_2}|B(t_i-\frac{k_2\pi}{\nu}).$$

It follows from (3.6) and the assumption $\lambda \geq N^{-1}$, that

$$\frac{1}{n_k+k^2\lambda} \leq C\max\left(\frac{1}{N}, \frac{1}{N^2\lambda}\right) \leq \frac{C}{N}, \text{ for all } k. \tag{A.14}$$

Since $(1/\nu)\sum_k B(k\pi/\nu+k'\pi/\nu) = (1/\nu)\sum_k B(k\pi/\nu) < \infty$, for all k', it follows from (A.14) that

$$\sum_{k=0}^{K} |b_k|B(\frac{k\pi}{\nu}+\frac{k'\pi}{\nu}) = \frac{1}{\nu}\sum_k \frac{1}{n_k+k^2\lambda}B(\frac{k\pi}{\nu}+\frac{k'\pi}{\nu}) \leq \frac{C}{N}. \tag{A.15}$$

Using this and the assumption that B is bounded,

$$T_1 \leq \frac{C}{N}\sum_{k_1=0}^{K}\frac{n_k}{n_k+k^2\lambda} \leq \frac{C}{N}\sum_{k_1=0}^{K}\frac{N}{N+k^2\lambda}.$$

Now $\sum_{k_1=0}^{K}\frac{N}{N+k^2\lambda} \leq 1+I$, where

$$I \equiv \int_0^\infty \frac{N}{N+\lambda x^2}dx = \sqrt{\frac{N}{\lambda}}\int_0^\infty \frac{1}{1+v^2}dv = \sqrt{\frac{N}{\lambda}}\frac{\pi}{2}.$$

Since $\lambda \leq N$, we conclude that $T_1 \leq C/\sqrt{N\lambda}$. The same can be concluded for T_2 and T_3 using similar derivations, and so $\mathrm{var}(\hat{f}_\lambda(\omega)) \leq C/\sqrt{N\lambda}$. This concludes the derivation for the bound of the variance.

We next prove the bound for the absolute bias. Clearly,

$$\begin{aligned}|\mathrm{bias}(\hat{f}_\lambda(\omega))| &= |f(\omega) - E[\hat{f}_\lambda(\omega)]| \\ &\le \Big|2\int_0^\infty \cos(u\pi\omega)R(u\pi)du - \frac{2}{\nu}\sum_{k=1}^{K}\frac{\cos(k\pi\omega/\nu)}{n_k+k^2\lambda}n_k R(\frac{k\pi}{\nu})\Big| + \frac{1}{\nu}\frac{1}{n_0}\mathbb{E}(S_0) \\ &\le U_1 + U_2 + U_3 + \frac{1}{\nu}\mathrm{var}(X(0)),\end{aligned}$$

where

$$\begin{aligned}U_1 &= 2\int_{K/\nu}^\infty |R(u\pi)|du, \\ U_2 &= \Big|\frac{2}{\nu}\sum_{k=1}^{K}\cos(\frac{k\pi\omega}{\nu})R(\frac{k\pi}{\nu}) - 2\int_0^{K/\nu}\cos(u\pi\omega)R(u\pi)du\Big|, \\ U_3 &= \frac{2}{\nu}\sum_{k=1}^{K}\frac{k^2\lambda}{n_k+k^2\lambda}|R(\frac{k\pi}{\nu})|.\end{aligned}$$

First, by (3.7) and (3.10),

$$U_1 \le C\int_{K/\nu}^\infty (u\pi)^{-\alpha-1}du \le C\Big(\frac{\nu}{K}\Big)^\alpha. \tag{A.16}$$

Now consider U_2. Letting $g(s) = \cos(s\pi\omega/\nu)R(s\pi/\nu)$, we obtain

$$U_2 \le \frac{1}{\nu}\Big|\sum_{k=1}^{K}g(k) - \int_0^K g(s)ds\Big| = \frac{1}{\nu}\Big|\int_0^K g(s)(dW(s)-ds)\Big|,$$

where $W(s) \equiv \sum_{k=1}^{K} I(k \le s)$. Using integration by parts and the fact that $\sup_s |W(s)-s| \le 1$, we obtain

$$\begin{aligned}\Big|\int_0^K g(s)(dW(s)-ds)\Big| &= \Big|\int_0^K g^{(1)}(s)(W(s)-s)ds\Big| \\ &\le \int_0^K |g^{(1)}(s)|ds \le \int_0^\infty |g^{(1)}(s)|ds,\end{aligned}$$

which is finite by condition (C2). Thus,

$$U_2 \le \frac{C}{\nu}. \tag{A.17}$$

Now consider U_3. By (3.7) and (3.10),

$$U_3 \le \frac{C}{\nu}\sum_{k=1}^{[\nu]}\frac{k^2\lambda}{n_k+k^2\lambda} + \frac{C}{\nu}\sum_{k=[\nu]+1}^{\infty}\frac{k^2\lambda}{n_k+k^2\lambda}(k/\nu)^{-\alpha-1} \equiv V_1 + V_2,$$

where $[\nu]$ denotes the integer part of ν. By (A.14),

$$V_1 = \frac{C}{\nu}\sum_{k=1}^{[\nu]} \frac{k^2\lambda}{n_k + k^2\lambda} \le \frac{C\lambda\nu^2}{N}. \tag{A.18}$$

Also, by (3.6),

$$V_2 = C\nu^\alpha\lambda \sum_{k=[\nu]+1}^{\infty} \frac{k^{1-\alpha}}{n_k + k^2\lambda} \le C\nu^\alpha\lambda\Big[\sum_{k=[\nu]+1}^{\infty} \frac{k^{1-\alpha}}{N + k^2\lambda} + \sum_{k=[\zeta N]}^{\infty} \frac{k^{1-\alpha}}{n_k + k^2\lambda}\Big].$$

Observe that

$$\begin{aligned}
\sum_{k=[\nu]+1}^{\infty} \frac{k^{1-\alpha}}{N + k^2\lambda} &\le \frac{([\nu]+1)^{1-\alpha}}{N + ([\nu]+1)^2\lambda} + \int_\nu^\infty \frac{x^{1-\alpha}}{N + x^2\lambda}dx \\
&= \frac{([\nu]+1)^{1-\alpha}}{N + ([\nu]+1)^2\lambda} + \Big(\frac{N}{\lambda}\Big)^{(2-\alpha)/2} \int_{\nu\sqrt{\lambda/N}}^\infty \frac{u^{1-\alpha}}{N(1+u^2)}du \\
&\le \frac{\nu^{1-\alpha}}{N + \nu^2\lambda} + \Big(\frac{N}{\lambda}\Big)^{(2-\alpha)/2} \frac{1}{N}\int_{\nu\sqrt{\lambda/N}}^\infty u^{1-\alpha}du \\
&= \frac{\nu^{1-\alpha}}{N + \nu^2\lambda} + \frac{1}{\alpha-2}\frac{1}{N}\nu^{2-\alpha}.
\end{aligned}$$

Since $\nu \ge \nu_0$, it is easy to see that the two terms on the right-hand side can be combined to give

$$\sum_{k=[\nu]+1}^{\infty} \frac{k^{1-\alpha}}{N + k^2\lambda} \le C\frac{\nu^{2-\alpha}}{N}.$$

Also,

$$\sum_{k=[\zeta N]}^{\infty} \frac{k^{1-\alpha}}{n_k + k^2\lambda} \le \frac{1}{\lambda}\sum_{k=[\zeta N]}^{\infty} k^{-1-\alpha} \le \frac{C}{N^\alpha}.$$

Thus, we obtain

$$V_2 \le C\left[\frac{\lambda\nu^2}{N} + \Big(\frac{\nu}{N}\Big)^\alpha\right]. \tag{A.19}$$

Summarizing the results from (A.16)-(A.19), we conclude that

$$|\mathrm{bias}(\hat{f}_\lambda(\omega))| \le C\left[\Big(\frac{\nu}{K}\Big)^\alpha + \frac{1}{\nu} + \frac{\lambda\nu^2}{N} + \Big(\frac{\nu}{N}\Big)^\alpha\right] \le C\left[\frac{1}{\nu} + \frac{\lambda\nu^2}{N} + \Big(\frac{\nu}{N}\Big)^\alpha\right],$$

since $N \le K$. This completes the proof of Theorem 3.

Proof of Theorem 5. As in the proof of Theorem 3, we have

$$\mathrm{var}(\tilde{f}_\lambda(\omega)) = \sum_{k_1=0}^{K}\sum_{k_2=0}^{K} b_{k_1}b_{k_2}\mathrm{cov}(S'_{k_1}, S'_{k_2}).$$

For $1 \le k \le K$, define $w_k(i,j) \equiv \mathbb{I}((i,j) \in L_k), 1 \le i \le j \le N$, and let $\mathcal{I}_k \equiv \{1 \le i \le N : \sum_{j=i}^{N} w_k(i,j) > 0\}$. It follows that

$$S'_k = \sum_{i\in\mathcal{I}_k}\sum_{j=i}^{N} w_k(i,j)X(t_i)X(t_j).$$

Following the proof of Theorem 3, we define $g(u;i,k) \equiv \sum_{j=i}^{N} w_k(i,j)a(t_j - u)$. Then, for $t_i \in (k_i\pi/\nu - \pi/2\nu, k_i\pi/\nu + \pi/2\nu]$,

$$|g(u;i,k)| \le \sup_{|\delta|\le\pi/\nu} |a(\frac{k_i\pi}{\nu} + \frac{k\pi}{\nu} - u + \delta)| \le \rho_\nu\beta(\frac{k_i\pi}{\nu} + \frac{k\pi}{\nu} - u).$$

The rest of the derivation of the variance inequality is almost identical to that of Theorem 3 using (C1′), and it is omitted.

We next prove the bound for the absolute bias. First,

$$\begin{aligned}|\mathrm{bias}(\tilde{f}_\lambda(\omega))| &= |f(\omega) - \mathbb{E}[\tilde{f}_\lambda(\omega)]| \\ &\le \Big|2\int_0^\infty \cos(u\pi\omega)R(u\pi)du - \frac{2}{\nu}\sum_{k=1}^{K}\frac{\cos(k\pi\omega/\nu)}{n'_k + k^2\lambda}\mathbb{E}(S'_k)\Big| + \frac{1}{\nu}\frac{1}{n'_0}\mathbb{E}(S'_0),\end{aligned}$$

where $\mathbb{E}(S'_k) = \sum_{(t_i,t_j)\in L_k}\mathbb{E}(X(t_i)X(t_j)) = \sum_{(t_i,t_j)\in L_k} R(|t_i - t_j|)$. Note that $|t_i - t_j| \in (k\pi/\nu - \pi/\nu, k\pi/\nu + \pi/\nu]$. By a Taylor expansion,

$$R(|t_i - t_j|) = R(\frac{k\pi}{\nu}) + R^{(1)}(\xi_{i,j,k})(|t_i - t_j| - \frac{k\pi}{\nu}),$$

where $\xi_{i,j,k} \in k\pi/\nu \pm \pi/\nu$. Mimicking the derivation of the bias in Theorem 3, write

$$|\mathrm{bias}(\tilde{f}_\lambda(\omega))| \le U_1 + U_2 + U_3 + U_4 + \frac{1}{\nu}\frac{1}{n'_0}\mathbb{E}(S'_0),$$

where U_1, U_2, U_3 are defined as in Theorem 3, except n_k there is replaced by n'_k here, and

$$U_4 \equiv \Big|\frac{2}{\nu}\sum_{k=1}^{K}\frac{\cos(k\pi\omega/\nu)}{n'_k + k^2\lambda}\sum_{(t_i,t_j)\in L_k} R'(\xi_{i,j,k})(|t_i - t_j| - \frac{k\pi}{\nu})\Big|.$$

The terms U_1, U_2, U_3 are dealt with as in Theorem 3, and the same bounds can be obtained. By (C2′),

$$\begin{aligned} U_4 + \frac{1}{\nu}\frac{1}{n_0'}\mathbb{E}(S_0') &\le \frac{C}{\nu^2}\sum_{k=0}^{K}\frac{n_k'}{n_k' + k^2\lambda}Q(\frac{k\pi}{\nu}) \\ &= \frac{C}{\nu^2}\sum_{k<\nu}\frac{n_k'}{n_k' + k^2\lambda} + \frac{C}{\nu^2}\sum_{k>\nu}\frac{n_k'}{n_k' + k^2\lambda}Q(\frac{k\pi}{\nu}) \\ &= O(\frac{1}{\nu}). \end{aligned}$$

Thus, the bound of $|\text{bias}(\tilde{f}_\lambda(\omega))|$ is derived.

References

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* **37**, 1-16.

Beltrão, K. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* **8**, 21-38.

Broersen, P. (2008). Time series models for spectral analysis of irregular data far beyond the mean data rate. *Measurement Sci. Tech.* **19**, 1-14.

Broersen, P. and Bos, R. (2006). Estimating time-series models from irregularly spaced data. *IEEE Trans. Instrumentation and Measurement* **55**, 1124-1131.

Bos, R., de Waele, S. and Broersen, P. (2002). Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. *IEEE Trans. Instrumentation and Measurement* **51**, 1289-1294.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Second Edition. Springer, New York.

Cressie, N. (1993). *Statistics for Spatial Data*. rev. edn. Wiley, New York.

Eyer, L. and Bartholdi, P. (1999). Variable stars: which Nyquist frequency? *Astronom. Astrophys. Supplement Series* **135**, 1-3.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B* **57**, 371-394.

Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25**, 359-369.

Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *Ann. Math. Statist.* **24**, 537-539.

Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*. Chelsea, New York.

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

Hurvich, C. and Beltrão, K. (1990). Cross-validatory choice of a spectrum estimate and its connection with AIC. *J. Time Ser. Anal.* **11**, 121-137.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral Analysis and Its Applications*. Holden-Day, San Francisco.

Kohn, R., Ansley, C. F. and Wong, C.-M. (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika* **79**, 335-346.

Li, K.-C. (1985). From Stein's unbiased risk estimates to the method of generalized cross-validation. *Ann. Statist.* **13**, 1352-1377.

Li, K.-C. (1987). Asymptotic optimality of C_p, C_L, cross-validation and generalized cross-validation, discrete index set. *Ann. Math. Statist.* **15**, 958-975.

Lomb, N. R. (1976). Least squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39**, 447-462.

Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Probab.* **5**, 439-468.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.* **1**, 502-527.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28**, 329-348.

Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle Likelihood. *J. Amer. Statist. Assoc.* **89**, 600-610.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. 1. Academic Press, London.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, URL `http://www.R-project.org.`

Scargle, J. D. (1992). Studies in astronomical time series analsysis II. Statistical aspects of spectral analsyis of unevenly spaced data. *Astrophys. J.* **263**, 835-853.

Stoica, P. and Sandgren, N. (2006). Spectral analysis of irregularly-sampled data; Paralleling the regularly-sampled data approaches. *Digital Signal Process.* **16**, 712-734

Tikhonov, A. N. and Arsenin, V. A. (1977). *Solution of Ill-posed Problems*. Winston & Sons, Washington.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.* **93**, 341-348.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*. Springer, New York.

Department of Statistics, Indiana University, Bloomington IN 47408, U.S.A.

E-mail: huang48@indiana.edu

Department of Statistics, University of Michigan, Ann Arbor MI 48109, U.S.A.

E-mail: thsing@umich.edu

Department of Statistics, The Ohio State University, Columbus OH 43210, U.S.A.

E-mail: ncressie@stat.osu.edu

(Received July 2090; accepted April 2010)