Spectral Density Estimation through a Regularized Inverse Problem

Chunfeng Huang[a], Tailen Hsing[b], and Noel Cressie[c]

October 30, 2008

Technical Report 08-06

Department of Statistics

Indiana University

Bloomington, IN 47405

[a]Departments of Statistics, Indiana University. huang48@indiana.edu
[b]Department of Statistics, University of Michigan. thsing@umich.edu
[c]Departments of Statistics, The Ohio State University. ncressie@stat.osu.edu

Spectral Density Estimation through a Regularized Inverse Problem

Summary. In the study of stationary processes on the real line, the spectral density function is a parameter of considerable interest. In this paper, we consider a new estimator of the spectral density function obtained by a regularized inversion of estimated covariances. In particular, the data are not required to be observed on a grid and the estimator is not based on the periodogram. For data that are observed on a grid, the estimator is derived in closed from, and the mean squared error of the estimator can be computed. A numerical study is also included to illustrate the methodology.

Running title: Spectrum estimation through a regularized inverse problem

Key words: Asymptotic theory, 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

$$R(t) = \int e^{it\omega} f(\omega) d\omega = 2 \int_0^\infty \cos(t\omega) f(\omega) d\omega, \tag{1}$$

and

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

The estimation of f is one of the oldest and most-studied statistics problems for stochastic processes. Suppose that the data $\{X(t), 0 \le t \le T\}$ are available. Define the periodogram

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

where

$$\hat{R}(t) = \frac{1}{T} \int_0^{T-|t|} X(u) X(u + |t|) du.$$

By (2), $I_T(\omega)$ is asymptotically unbiased for $2\pi f(\omega)$, but not consistent in that the variance does not tend to zero as $T \to \infty$. However, 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-parameter processes (i.e., time series) have received more attention than continuous-parameter processes or spatial processes. For early literature, see Bartlett (1950), Grenander and Rosenblatt (1953), Parzen (1957), Jenkins and Watts (1968), and Priestley (1981). More recent works on data-driven procedures based on the periodogram include Hurvich (1985), Beltrão and Bloomfield (1987), Hurvich and Beltrão (1990), Pawitan and O'Sullivan (1994), and Fan and Kreutzberger (1998), to name a few.

In this paper we consider an approach for estimating f without directly using (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, each of which entailing considerations unique to that setting. In this paper, we will focus on stationary processes indexed by $\mathbb{R}$. We believe that this is a good first step in understanding the nature of this new approach. In Section 2, we define the notation and describe the basic methodology that leads to the new spectral density estimator. In Section 3, we consider gridded data and derive a representation of the estimator for that case. A special case of gridded data is time series data. While we emphasize that our method should not be viewed as a purely stationary time-series method, comparisons can be made with other time-series methods in that special context. In Section 4, we describe a weighted cross-validation approach for choosing the smoothing parameter in our method, and compare that procedure with a number of bench-mark spectral procedures for stationary time series. Moreover, in Section 4, we demonstrate numerically how our method can be applied to continuous-parameter stationary processes for which the data are observed at random locations. Section 5 considers the computation of bounds of the mean squared error for gridded data, and discussions and conclusions are given in Section 6. All of the proofs are given in the Appendix.

2. Methodology. In the developments below, it will be convenient notationally to absorb the constant 2 into f in equation (1); that is, from now on we will write

$$R(t) = \int_0^\infty \cos(t\omega) f(\omega) d\omega. \tag{3}$$

Suppose that we observe the process $X(t)$ at $t = t_i, 1 \leq i \leq N$. 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 will be small for a function g close to f:

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

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 (cf. O'Sullivan, 1986), for which regularization is essential.

A computationally efficient approach to address this problem is to use the Sobolev-space setting. 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 $n = N + \binom{N}{2}$; and let $h_i, 1 \le i \le n$, be the corresponding differences, $|t_j - t_k|$, between the observational time points for the pairs. Due to the aliasing effect, instead of integration on $[0, \infty)$ as in (3), let $[0, \nu]$ be the support of the spectral-density estimator, $\nu > 0$. Define

$$h'_i \equiv \nu h_i, \ 1 \le i \le n, \tag{4}$$

and the bounded linear functional,

$$L_i g \equiv \int_0^1 \cos(h'_i u) g(u) du, \ g \in W_1 \equiv W_1[0,1],$$

where $W_1[0,1]$ is the Sobolev-Hilbert space of order 1 containing measurable functions on $[0,1]$ that are absolutely continuous with square-integrable derivatives. Write

$$W_1 = \mathcal{H}_0 \oplus \mathcal{H}_1,$$

where $\mathcal{H}_0$ is the space of constant functions on $[0,1]$, and $\mathcal{H}_1$ is the space of absolutely continuous, measurable functions f on $[0,1]$ with $f(0) = 0$ and $\int_0^1 [f'(u)]^2 du < \infty$.

Denote by $\mathcal{P}_1$ the projection operator onto $\mathcal{H}_1$ in W_1. Readers are referred to Wahba (1990) for details. The new estimator of f in (3) that we propose is based on the following regularized optimization problem. Find $g \in W_1$ to minimize

$$\sum_{i=1}^{n}(y_i - L_i g)^2 + \lambda \int_0^1 [g'(u)]^2 du, \tag{5}$$

where $\lambda > 0$ is a smoothing parameter. It is a crucial part of the methodology that g can be solved algebraically as follows. Let η_i be the representer for L_i, so that

$$\langle \eta_i, g \rangle = L_i g, \;\; g \in W_1,$$

where

$$\langle g_1, g_2 \rangle \equiv g_1(0)g_2(0) + \int_0^1 g_1'(u)g_2'(u)du, \;\; g_1, g_2 \in W_1, \tag{6}$$

and let

$$\xi_i = \mathcal{P}_1 \eta_i.$$

Recall that the reproducing kernel of W_1 is (Wahba, 1990, p. 8)

$$\mathcal{R}(s,t) = 1 + \min(s,t), \;\; s,t \in [0,1].$$

Let $\mathcal{R}_s = \mathcal{R}(s,\cdot)$. It follows that

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

and

$$\xi_i(s) = (P_1\eta_i)(s) = \begin{cases} s - \frac{1}{2}s^2 & h_i' = 0, \\ \frac{\sin(h_i')h_i's + \cos(h_i's) - 1}{h_i'^2} & h_i' \neq 0. \end{cases} \tag{7}$$

Define the vectors

$$\mathbf{y} \equiv (y_1, \ldots, y_n)^T,$$

$$T_{n\times 1} \equiv \begin{bmatrix} L_1 1 \\ L_2 1 \\ \vdots \\ L_n 1 \end{bmatrix} = \begin{bmatrix} \int_0^1 \cos(h_1' u) du \\ \int_0^1 (\cos(h_2' u) du \\ \vdots \\ \int_0^1 \cos(h_n' u) du \end{bmatrix}, \tag{8}$$

where "1" is used to represent the constant function equal to unity, and define the matrices

$$\Sigma_{n\times n} \equiv \{\langle \xi_i, \xi_j \rangle\} \quad \text{and} \quad M_{n\times n} \equiv \Sigma + \lambda I. \tag{9}$$

Then it follows from Theorem 1.3.1 of Wahba (1990) that the solution of the optimization problem in (5) is given by

$$\hat{g}_\lambda(u) \;=\; d + \sum_{i=1}^{n} c_i \xi_i(u), \;\; u \in [0, 1], \tag{10}$$

where

$$d = (T^T M^{-1} T)^{-1} T^T M^{-1} \mathbf{y}, \tag{11}$$

and

$$\mathbf{c} = M^{-1}(I - T(T^T M^{-1} T)^{-1} T^T M^{-1})\mathbf{y}. \tag{12}$$

Finally, let

$$\hat{f}_\lambda(\omega) = \begin{cases} \frac{1}{\nu}\hat{g}_\lambda(\omega/\nu), & \omega \in [0, \nu], \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

In the rest of this paper, we will focus on the properties of this estimator.

Note that the algorithm in general does not require the data to be observed on a grid, although gridded data do lead to simpler formulas for the estimator's bias and variance (see Section 5). The algorithm can also be extended to processes whose index sets are multi-dimensional by working with more general Sobolev spaces.

It is important to note that $\hat{f}_\lambda(\omega)$ could be negative. However, in all of the examples that we have tried, this is a minor issue. We will demonstrate this point numerically by estimating a spectrum which is partly equal to 0 in section 4. Section 9.4 of Wahba (1990) describes a simple way to deal with this problem. Let $u_1, \ldots, u_\ell$ be a set of points in $[0, 1]$. Adding the constraints $g(u_1), \ldots, g(u_\ell) \geq 0$ in the optimization problem (5), the solution can be written as $\check{f}_\lambda(\omega) = (1/\nu)\check{g}_\lambda(\omega/\nu)$, $\omega \in [0, \nu]$, with

$$\check{g}_\lambda = \sum_k b_k R_1(u_k, \cdot) + d + \sum_{i=1}^n c_i \xi_i,$$

where R_1 is the reproducing kernel of $\mathcal{H}_1$, and the constants b_k, c_i, d can be obtained by solving a quadratic programming problem. Since the functions in $\mathcal{H}$ 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 optimally choose smoothing parameters, and the absence of closed-form solutions that could be useful for theoretical considerations. The latter two points will be amplified in future sections. To fundamentally address the lack of complete nonnegativity, one approach is to consider the estimation of the logarithm of f. This is common in time series analysis, where the fact that the periodogram 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. Both of these approaches will be explored in future work.

While we emphasize a continuous-parameter stationary process in this paper, the algorithm applies readily to a discrete-parameter stationary process. If $X(t)$ is observed at $t = 1, 2, \ldots, N$, the natural choice for the support parameter ν is π (see next paragraph), in which case the spectral density will be estimated on the interval $[0, \pi]$.

The choices of ν and λ in (4) and (5) are obviously crucial. We defer the discussion on the choice of λ until Section 4. A practical consideration in choosing ν is avoiding the aliasing effect. For example, if the t_i are equally spaced with spacing τ, then, by Nyquist's Sampling Theorem, choose $\nu/2\pi$ (highest identifiable frequency) $< 1/2\tau$ (one half the sampling rate). That is, $\nu < \pi/\tau$. See Priestley (1981).

3. Gridded data. In this section, assume that the observational points $t_1 < \ldots < t_N$ 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 consecutively observed on $\mathcal{G}$, and hence our results are relevant for a spatial process in one dimension and should not be viewed as method only for the usual time series.

Define

$$K \equiv \frac{(t_N - t_1)\nu}{\pi},$$

$$n_k \equiv \sum_{i=1}^{N} \sum_{j=i}^{N} I(t_j - t_i = k\pi/\nu), \;\; k = 0, 1, 2, \ldots, K,$$

and

$$n \equiv \sum_{k=0}^{K} n_k.$$

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 k-th sub-sequence is not an issue. 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; the next n_2 y_i's are products of pairs of data that are observed at distance $2\pi/\nu$ 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 in (13). The following can be proved:

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, namely $t_i = i$. The periodogram can be written as

$$I_T(\omega) = \frac{1}{N} S_0 + \frac{2}{N} \sum_{k=1}^{K} \cos(k\omega) S_k.$$

The natural choice of ν in our procedure is $\nu = \pi$, in which case the difference between $\hat{f}_\lambda$ and $\pi^{-1} I_T$ is that in the summation over $1 \leq k \leq K$, the coefficient of $\cos(k\omega)S_k$

is $1/[n_k + 2(k\pi)^2\lambda]$ for $\hat{f}_\lambda$, where $n_k = N - k$, instead of $1/N$ for $\pi^{-1}I_T$. Intuitively, having n_k in the denominator reduces the bias since that is the number of terms in S_k. While the other component $2(k\pi)^2\lambda$ in the denominator seems to make the bias worse for $\hat{f}_\lambda$ as k increases, its real effect is down-weighting 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. See pp. 432-449 of Priestley (1981) and pp. 351-382 of Brockwell and Davis (1991). However, the weights in the two procedures are different in form, and, more importantly, the weights in our estimator 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

$$\sum_{k=1}^{K} n_k \left(S_k/n_k - \int_0^1 \cos(k\pi u) g(u) du \right)^2 + \lambda \int_0^1 [g'(u)]^2 du, \tag{14}$$

and then follow the same steps as (5) through (13) to obtain the estimator $\tilde{f}_\lambda$ of f.

Corollary 2. *The estimator $\tilde{f}_\lambda$ defined by (14) is identically equal to $\hat{f}_\lambda$ given by Theorem* 1 for all $\lambda > 0$.

4. Cross-validation and numerical results. This section contains a discussion on how to choose the smoothing parameter from data, comparisons with some time-series procedures, and an example of how to implement our procedure when the data are observed from a continuous-parameter process.

Choosing λ by 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

$$GCV(\lambda) \equiv \frac{\|(I - H_\lambda)\mathbf{y}\|^2}{[\mathrm{tr}(I - H_\lambda)]^2}, \ \lambda > 0, \tag{15}$$

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

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

and T and M are defined by (8) and (9) in Section 2. One would then choose λ as the minimizer of $GCV(\lambda)$. However, as pointed out on p. 65 of Wahba (1990), GCV is likely to give unsatisfactory results when $\{y_i\}$ are highly correlated. This turns out to be the case due to the way we formulated the problem.

There has not been much work on GCV for dependent data; see Wang (1998) and the references therein. Here we consider two possibilities. Consider the loss $L(\lambda) = (1/n)||\hat{y} - E(\mathbf{y})||^2$ where $\hat{y} = H_\lambda \mathbf{y}$. An unbiased estimator of the risk $\mathrm{E}L(\lambda)$ is

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

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

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

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

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

One can then choose λ to minimize

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

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 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,

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

where Ξ_k is the k-th $n_k \times n_k$ diagonal block matrix of Ξ. 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

$$f(\omega) = I(0 \le \omega < .4) + \frac{1}{2}\left[1 + \cos\left(\frac{\pi(\omega - .4)}{\pi - .8}\right)\right] I(.4 \le \omega \le \pi - .4),\ \omega \ge 0. \quad (16)$$

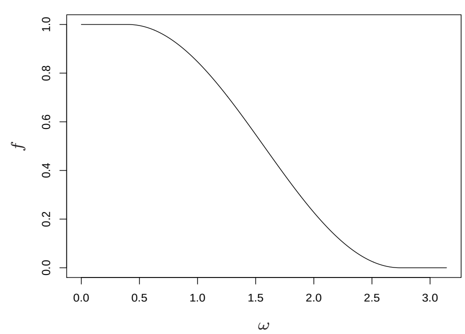

Figure 1: *The spectral density f in (16)*

Let the process be observed at $t = 1, 2, \ldots, 2000$. To illustrate our method, for $\lambda_i = 10^{u_i}$, where $u_i = -5 + 14(i-1)/49, 1 \le i \le 50$, we computed $L(\lambda_i)$, $V(\lambda)$ and $GCV'(\lambda_i)$ based on one simulation run. These three criteria versus i are shown in Figure 2. It can be seen that the optimal λ determined by these three functions are

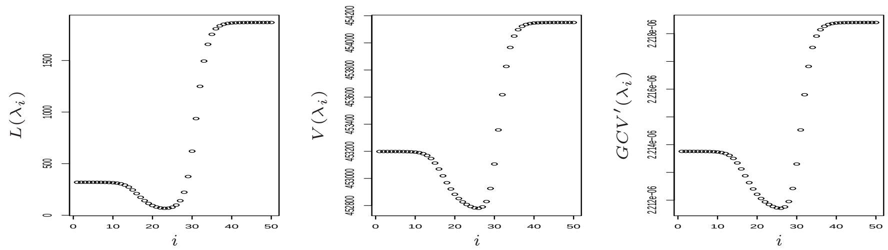

Figure 2: *The left-most plot is $L(\lambda_i)$ versus i; the middle plot is $V(\lambda_i)$ versus i; the right plot is $GCV'(\lambda_i)$ versus i.*

remarkably close, and $V(\lambda)$ and $GCV'(\lambda)$ are very similar to each other. We use $GCV'(\lambda)$ in the following comparisons.

Comparisons with time-series procedures

While our procedure is not restricted to time series, a comparison with leading spectrum-estimation procedures in the time-series context will be insightful. The procedures chosen here are the following well-known, data-driven procedures:

Method 1: 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).

Method 2: The local linear smoother $\hat{m}_{\rm LS}$ and local maximum likelihood estimator $\hat{m}_{\rm LK}$ introduced in Fan and Kreutzberger (1998), with the smoothing parameters selected by the constant bandwidth selector in Fan and Gijbels (1995).

Method 3: The nonparametric penalized Whittle likelihood approach due to Pawitan and O'Sullivan (1994).

We made the comparisons by simulations using a collections of time-series models.

It was found that our procedure competes well with the other procedures in general. Here, again, we focus on the stationary Gaussian time series with spectral density f given by (16). Four hundred simulation runs were performed, where, for each run, the spectrum estimate $\hat{f}$ was computed for each of the five methods based on data $X(t), t = 1, 2, \ldots, 2000$. 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 3. The abbreviations "HHC", "BB", "PS", "FK.LS", and "FK.LK" stand for, respectively, our estimator, the Beltrão and Bloomfield estimator, the Pawitan and O'Sullivan estimator, and Fan and Kreutzberger's $\hat{m}_{\mathrm{LS}}$ and $\hat{m}_{\mathrm{LK}}$. It can been seen that our estimator performed quite a bit better than the Beltrão and Bloomfield estimator, somewhat better than the Fan and Kreutzberger estimators, and was comparable to the Pawitan and O'Sullivan estimator.

We now revisit the issue that our estimator $\hat{f}$ may potentially take on negative values. Note that the spectrum f used in this example is equal to zero on $[\pi - 1/4, \pi]$. Thus, if the negativity of $\hat{f}$ is a prevalent issue in this methodology, we would expect to see a substantial number of negative values in $\hat{f}(\omega)$ for ω close to π. In fact, $\hat{f}$ estimated f exceedingly well close to π, but only 18 out of 400 runs produced estimates that are not completely positive. The percentage of negative values in all of the estimates out of all of the runs was roughly .6%, and the minimum of the values was -0.050.

A final point for this subsection is that, in conducting the comparisons, we found that our procedure is considerably easier to code than the other procedures. Our procedure is also by far the most computationally efficient, which only required a small fraction of computing time required by the Pawitan and O'Sullivan and the Fan and Kreutzberger procedures in the simulations.

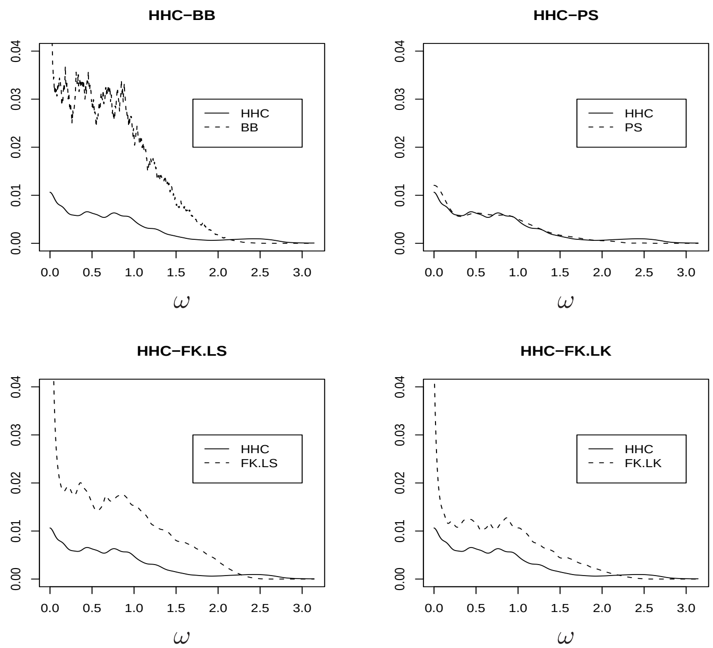

Figure 3: *Mean squared error comparisons of our estimator with other estimators. The upper-left plot compares with the Beltrão and Bloomfield method, the upper-right plot compares with the Pawitan and O'Sullivan estimator, and the two lower plots compare with the estimators of Fan and Kreutzberger.*

<u>Continuous-parameter process and covariance estimation</u>

We have demonstrated that our methodology competes well with periodogram-based approaches in the context of time series. However, one enormous advantage of our approach is the flexibility it offers in terms of the wide range of spectral estimation problems that can be readily addressed. As a simple demonstration, we considered 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 2000 points which are iid uniformly distributed on $[0, 1000]$. Since the dimension of M is huge, inverting it is computationally costly. To find an approximate solution, we let $\nu = 6\pi$ and replaced each time point by the nearest grid point $k/6, k = 0, 1, 2 \ldots$. We then carried out the analysis using the computational formulas derived for gridded data in Theorem 1, with the smoothing parameter selected by GCV_1. Note that in this case, we could have multiple observations or no observation at all at any grid point. To our knowledge, there is no counterpart methodology available using the periodogram approach. The choice of ν clearly mattered, but was not so crucial. In addition to $\nu = 6\pi$, we also tried $\nu = k\pi$ for k between 7 and 10 and the outcomes were similar.

We present the outcome of one simulation run in Figure 4, where the plot contains the true (dotted line) and estimated spectrum (solid line) on $[0, 5]$. Outcomes from other runs are qualitatively similar.

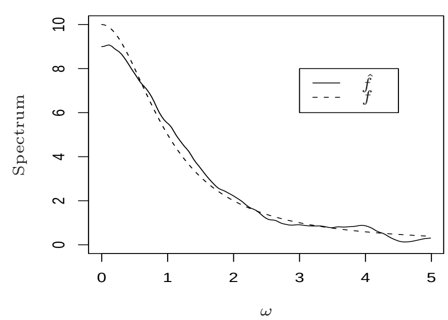

Figure 4: *Estimation of spectrum of a continuous-parameter process.*

5. The mean squared error of the estimator. We assume in this section that the data are observed on a grid, and we continue to use the notation developed in Section 3. Our goal is compute the bias and variance bounds for $\hat{f}_\lambda(\omega)$. Note that the notation C will be a generic symbol for a finite positive constant whose value may be different in different places. Also, proofs of the results are found in the Appendix.

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),

$$X(t) = \int a(t-s)dZ(s), \ t \in \mathbb{R}, \tag{17}$$

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, which will simplify the derivations considerably. Let

$$E[Z(dt)]^2 = dt \quad \text{and} \quad E[Z(dt)]^4 = \mu_4 dt, \tag{18}$$

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

Also assume that the observational points $t_i, 1 \le i \le N$, are such that for some $\zeta, \delta \in (0,1)$,

$$\inf_{k \le \zeta N} n_k \ge \delta N. \tag{19}$$

Recall that we do not require data to be consecutively observed on the grid $\mathcal{G}$. The condition (19) insures 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 consecutively observed on $\mathcal{G}$.

The following regularity conditions are also needed.

(C1) Let β be a nonnegative measurable function on $\mathbb{R}$, and B a bounded, symmetric, integrable function on $\mathbb{R}$ with $B(t) \downarrow$ for $t > 0$. Recall the function a in (17).

Then assume that

$$|a(u)| \leq \beta(u), \text{ for all } u, \tag{20}$$

$$\int \beta(u)\beta(u+t)du \leq B(t), \text{ for all } t, \tag{21}$$

and for some $\nu_0 > 0$,

$$\sup_{\nu \geq \nu_0} \frac{1}{\nu} \sum_k \beta(k\pi/\nu + u)\beta(k\pi/\nu + u + t) \leq B(t), \text{ for all } u, t. \tag{22}$$

(C2) There exists $\alpha > 2$ and $C > 0$ such that the function B in (C1) satisfies

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

(C3) The covariance function $R(t) \equiv E[X(0)X(t)]$ is differentiable with $\int |R'(t)|dt < \infty$.

The condition (C1) has the following consequence. Since, by (18),

$$R(t) = \int a(u)a(u+t)dt,$$

the conditions (20) and (21) imply that

$$|R(t)| \leq B(t), \quad \text{for all } t. \tag{23}$$

The assumption that B is integrable therefore implies that X is a short-memory process (cf. Brockwell and Davis, 1991, Section 13.2).

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

THEOREM 3. *Assume that (C1) holds. Then there exists a bounded universal constant C such that for all $\nu \geq \nu_0$, $\omega \in [0, \nu]$, N satisfying (19), and $\lambda \in [N^{-1}, N]$, we have*

$$\text{var}(\hat{f}_\lambda(\omega)) \leq \frac{C}{\sqrt{N\lambda}}. \tag{24}$$

If, additionally, (C2) and (C3) hold, then there exists a bounded universal constant C such that for all ω, N, ν, λ specified above,

$$|bias(\hat{f}_\lambda(\omega))| \le C \left[\frac{1}{\nu} + \frac{\lambda\nu^2}{N} + \left(\frac{\nu}{N}\right)^\alpha \right]. \tag{25}$$

Corollary 4. *Assume that (C1)-(C3) hold. Then letting*

$$\lambda = \frac{N^{3/5}}{\nu^{8/5}}, \tag{26}$$

there exists a bounded universal constant C such that for all $\omega \in [0, \nu]$, N satisfying (19), and ν satisfying $N \ge \nu \ge \nu_0$,

$$MSE(\hat{f}_\lambda(\omega)) \le C \left[\left(\frac{\nu}{N}\right)^{4/5} + \frac{1}{\nu^2} \right]. \tag{27}$$

Proof. The proof follows immediately from Theorem 3. □

Note that $N\pi/\nu$ is roughly the range of the data. As such, (27) can be interpreted as

$$\mathrm{MSE}(\hat{f}_\lambda(\omega)) \le C \left[\frac{1}{(\text{range of data})^{4/5}} + \frac{1}{\nu^2} \right].$$

Suppose $X(t)$ is continuously observed for $t \in [0, T]$. Then, taking $N = [T\nu]$, where ν satisfies $\nu^{-2} \le T^{-4/5}$, and λ given by (26), we obtain

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

Theorem 3 can be generalized in a number of ways, including relaxing (19) and condition (C2), and not restricting the observation points to a grid. These extension, while useful, will make the proofs longer and more technical. We feel that the present set of conditions strike a balance between generality and ease of presentation.

While Theorem 3 and Corollary 4 are proved for the continuous-parameter process (17), a quick inspection of the proofs reveals that they also hold for the discrete-parameter process under parallel assumptions. 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 under regularity conditions; see the discussions on pp. 567-568 of Priestley (1981) and Section 4.7 of Grenander and Rosenblatt (1984).

6. Discussion and conclusions. We restricted our estimator to be in the Sobolev space W_1 to minimize technical difficulties in our derivations and the requirement on the smoothness of the spectral density. It would be interesting to consider the properties of the estimator when the space is taken to be W_k, the Sobolev space of order k. Also, in this paper, we limited our attention to a stationary process on $\mathbb{R}$. The formulation of our methodology can, in principle, be adapted for spectrum estimation of spatial processes that are stationary or are intrinsic random functions. These extensions will be investigated in future work.

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

(i) does not require the data to be observed on a regular grid, and holds strong promise in being adapted to more general spectral analysis settings such as the intrinsic random functions;

(ii) is computationally efficient;

(iii) has a fast rate of convergence;

(iv) does not make use of the periodogram in the time-series setting where it performs in general comparably to, and sometimes better than bench-mark periodogram-based procedures.

References

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

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

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

[4] 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.

[5] Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scandinavian Journal of Statistics* **25**, 359-369.

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

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

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

[9] 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.

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

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

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

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

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

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

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

[17] Priestley, M. B. (1981). *Spectral Analysis and Time Series*, vol. 1. Academic Press, London.

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

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

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

Appendix

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:

$$B_k \equiv \left[\frac{1}{2[(k-1)\pi]^2} J_{n_{k-1} \times n_{k-1}} + \lambda I_{n_{k-1}}\right]^{-1}, \ 2 \le k \le K+1, \tag{28}$$

$$A_1 \equiv I_{n_0}, \ A_k \equiv \frac{1}{[(k-1)\pi]^2} B_k J_{n_{k-1} \times n_0}, \ 2 \le k \le K+1, \tag{29}$$

$$A_0 \equiv \left[\frac{1}{3} J_{n_0 \times n_0} + \lambda I_{n_0}\right] - \sum_{k=2}^{K+1} \frac{1}{[(k-1)\pi]^2} J_{n_0 \times n_{k-1}} A_k, \tag{30}$$

and write

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

First note that by the identity

$$\left(\frac{1}{c} J_{k \times k} + I_k\right)^{-1} = I_k - \frac{1}{k+c} J_{k \times k}, \ c \ne 0, \tag{31}$$

we obtain

$$B_k = \lambda^{-1} \left(I_{n_{k-1}} - \frac{1}{n_{k-1} + 2[(k-1)\pi]^2 \lambda} J_{n_{k-1} \times n_{k-1}}\right), \tag{32}$$

which implies that

$$A_k = \frac{1}{[(k-1)\pi]^2} B_k J_{n_{k-1} \times n_0} = \frac{2}{n_{k-1} + [(k-1)\pi]^2 \lambda} J_{n_{k-1} \times n_0} \ . \tag{33}$$

LEMMA A1.

$$A_0^{-1} = \lambda^{-1} I_{n_0} - \lambda^{-1} \frac{1}{n_0 + \lambda/b_0} J_{n_0 \times n_0}, \tag{34}$$

where

$$b_0 \equiv \frac{1}{3} - 2 \sum_{k=2}^{K+1} \frac{1}{[(k-1)\pi]^2} \frac{n_{k-1}}{n_{k-1} + 2[(k-1)\pi]^2 \lambda} \ . \tag{35}$$

PROOF. By (30) and (33),

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

The result then follows from (31).

□

LEMMA A2.

$$M^{-1} = diag\{0, B_2, \ldots, B_{K+1}\} + A A_0^{-1} A^T.$$

PROOF. Recall the definition of Σ and M in (9). We first compute Σ. It follows from (7) that

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

Straightforward calculations using (6) show that

$$\langle \xi_i, \xi_j \rangle = \int_0^1 (1-s)^2 ds = \frac{1}{3}, \ \ h_i' = h_j' = 0,$$

$$\langle \xi_i, \xi_j \rangle = -\int_0^1 (1-s) \frac{\sin(k\pi s)}{k\pi} ds = -\frac{1}{(k\pi)^2}, \ \ h_i' = 0, h_j' = k\pi, k \neq 0,$$

$$\langle \xi_i, \xi_j \rangle = \int_0^1 \left(\frac{\sin(k\pi s)}{k\pi} \right)^2 ds = \frac{1}{2(k\pi)^2}, \; h_i' = h_j' = k\pi, k \neq 0,$$

$$\langle \xi_i, \xi_j \rangle = \int_0^1 \left(\frac{\sin(k_1 \pi s)}{k_1 \pi} \right) \left(\frac{\sin(k_2 \pi s)}{k_2 \pi} \right) ds = 0, \; h_i' = k_1\pi, h_j' = k_2\pi, 0 \neq k_1 \neq k_2 \neq 0.$$

It follows that

$$\Sigma = \begin{bmatrix} \frac{1}{3} J_{n_0 \times n_0} & -\frac{1}{\pi^2} J_{n_0 \times n_1} & -\frac{1}{(2\pi)^2} J_{n_0 \times n_2} & \cdots & -\frac{1}{(K\pi)^2} J_{n_0 \times n_K} \\ -\frac{1}{\pi^2} J_{n_1 \times n_0} & \frac{1}{2\pi^2} J_{n_1 \times n_1} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ -\frac{1}{(K\pi)^2} J_{n_K \times n_0} & 0 & 0 & \cdots & \frac{1}{2(K\pi)^2} J_{n_K \times n_K} \end{bmatrix}.$$

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

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

It is clear that

$$M^{k1} = A_k M^{11}, \; 2 \leq k \leq K+1, \quad \text{and} \quad M^{11} = A_0^{-1}.$$

Proceeding in this manner, in general we have

$$M^{kj} = A_k A_0^{-1} A_j^T, \quad k \neq j,$$

$$M^{11} = A_0^{-1},$$

$$M^{kk} = B_k + A_k A_0^{-1} A_k^T, \quad 2 \leq k \leq K+1.$$

□

By (8) and the assumptions on $\{h_i\}$, we have

$$T = \begin{bmatrix} \mathbf{1}_{n_0} \\ \mathbf{0}_{n-n_0} \end{bmatrix}. \tag{36}$$

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 (12),

$$\mathbf{c} = \lambda^{-1}(I - H_\lambda)\mathbf{y}. \tag{37}$$

LEMMA A3. *(i)*

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

where b_0 is given by (35).

(ii)

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

PROOF. Since $A_1 = I_{n_0}$, it follows from (36) and Lemma A2 that

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

Hence, (i) follows simply from Lemma A1 and (31).

Similarly, we also have

$$M^{-1}T(T^TM^{-1}T^T)^{-1}T^TM^{-1} = a_0AA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A^T.$$

By this and Lemma A2,

$$\begin{aligned} H_\lambda &= I - \lambda\ \mathrm{diag}\{0, B_2, \ldots, B_{K+1}\} - \lambda AA_0^{-1}A^T + \lambda a_0AA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A^T \\ &= \mathrm{diag}\{I_{n_0}, I_{n_1} - \lambda B_2, \ldots, I_{n_K} - \lambda B_{K+1}\} + \lambda\tilde{A}, \end{aligned} \tag{38}$$

where

$$\begin{aligned} \tilde{A} &= -AA_0^{-1}A^T + a_0AA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A^T \\ &= \{-A_iA_0^{-1}A_j^T + a_0A_iA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T\}_{i,j=1,\ldots,K+1}. \end{aligned}$$

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 (29),

$$\begin{aligned} & a_0A_iA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T \\ &= \frac{a_0}{[(i-1)\pi]^2}B_iJ_{n_{i-1}\times n_0}A_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T \\ &= \frac{a_0}{[(i-1)\pi]^2}B_i\mathbf{1}_{n_{i-1}}\mathbf{1}_{n_0}^TA_0^{-1}\mathbf{1}_{n_0}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T \\ &= \frac{a_0}{[(i-1)\pi]^2}B_i\mathbf{1}_{n_{i-1}}a_0^{-1}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T \\ &= \frac{1}{[(i-1)\pi]^2}B_i\mathbf{1}_{n_{i-1}}\mathbf{1}_{n_0}^TA_0^{-1}A_j^T = A_iA_0^{-1}A_j^T. \end{aligned}$$

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 (34) 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 $\tilde{A}$ =diag$\{-I_{n_0} + n_0^{-1}J_{n_0\times n_0}, 0, \cdots, 0\}$. Thus, (ii) follows from (38) and (32).

$\square$

By (i) of Lemma A3 and Lemma A1,

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

and, by (33),

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

It now follows from (11) that

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

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 (37), (ii) of Lemma A3, and (7),

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

The first term is equal to 0, whereas, by (32), the second term is equal to

$$
\begin{aligned}
&\sum_{k=1}^{K}\mathbf{1}_{n_k}^T\left(\lambda^{-1}I_{n_k}-\lambda^{-1}\frac{1}{n_k+2(k\pi)^2\lambda}J_{n_k\times n_k}\right)\mathbf{y}_k\frac{\cos(k\pi u)-1}{(k\pi)^2}\\
=\ &\sum_{k=1}^{K}\left(\lambda^{-1}S_k-\lambda^{-1}\frac{n_k}{n_k+2(k\pi)^2\lambda}S_k\right)\frac{\cos(k\pi u)-1}{(k\pi)^2}\\
=\ &2\sum_{k=1}^{K}\frac{(k\pi)^2}{n_k+2(k\pi)^2\lambda}S_k\frac{\cos(k\pi u)-1}{(k\pi)^2}\\
=\ &2\sum_{k=1}^{K}\frac{1}{n_k+2(k\pi)^2\lambda}S_k(\cos(k\pi u)-1).
\end{aligned}
$$

Combining d and $\sum_i c_i\xi_i(u)$ and making the transformation (13) gives the form of $\hat{f}_\lambda(\omega)$ in Theorem 1. This concludes the proof of Theorem 1 $\square$

Proof of Corollary 2:

The proof follows from a simple adaptation of that of Theorem 1. Redefine the first n_0 y_i's to be all equal to S_0/n_0, the next n_1 y_i's to be all equal to S_1/n_1, and so on.
$\square$

Proof of Theorem 3:

We reiterate the assumption 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

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

For $1 \le k \le K$, define

$$w_k(i,j) = I(t_j - t_i = k\pi/\nu),\ 1 \le i \le j \le N.$$

Let

$$I_k = \{1 \le i \le N : \sum_{j=i}^{N} w_k(i,j) > 0\},$$

so that

$$S_k = \sum_{i \in I_k}\sum_{j=i}^{N} w_k(i,j)X(t_i)X(t_j).$$

Then

$$\begin{aligned}
&\mathrm{cov}(S_{k_1}, S_{k_2}) \\
= &\sum_{i_1 \in I_{k_1}}\sum_{i_2 \in I_{k_2}}\sum_{j_1=i_1}^{N}\sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1)w_{k_2}(i_2,j_2)\mathrm{cov}(X(t_{i_1})X(t_{j_1}), X(t_{i_2})X(t_{j_2})) \\
= &\sum_{i_1 \in I_{k_1}}\sum_{i_2 \in I_{k_2}}\sum_{j_1=i_1}^{N}\sum_{j_2=i_2}^{N} w_{k_1}(i_1,j_1)w_{k_2}(i_2,j_2)\Big\{E[X(t_{i_1})X(t_{j_1})X(t_{i_2})X(t_{j_2})] - \\
&\qquad E[X(t_{i_1})X(t_{j_1})]E[X(t_{i_2})X(t_{j_2})]\Big\}
\end{aligned}$$

$$= \quad A_1 - A_2,$$

where

$$\begin{aligned} A_1 &= \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \sum_{j_1=i_1}^{N} \sum_{j_2=i_2}^{N} w_{k_1}(i_1, j_1) w_{k_2}(i_2, j_2) \int_{u_1} \int_{u_2} \int_{u_3} \int_{u_4} a(t_{i_1} - u_1) a(t_{j_1} - u_2) \\ &\qquad a(t_{i_2} - u_3) a(t_{j_2} - u_4) E[Z(du_1) Z(du_2) Z(du_3) Z(du_4)], \\ A_2 &= \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \sum_{j_1=i_1}^{N} \sum_{j_2=i_2}^{N} w_{k_1}(i_1, j_1) w_{k_2}(i_2, j_2) R(t_{i_1} - t_{j_1}) R(t_{i_2} - t_{j_2}). \end{aligned}$$

Write

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

By condition (C1),

$$|g(u; i, k)| \leq \beta(t_i + k\pi/\nu - u). \tag{39}$$

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

$$\begin{aligned} A_{11} &= \mu_4 \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \int_u a(t_{i_1} - u) g(u; i_1, k_1) a(t_{i_2} - u) g(u; i_2, k_2) du, \\ A_{12} &= \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \int_{u_1} \int_{u_2} a(t_{i_1} - u_1) g(u_1; i_1, k_1) a(t_{i_2} - u_2) g(u_2; i_2, k_2) du_1 du_2, \\ A_{13} &= \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \int_{u_1} \int_{u_2} a(t_{i_1} - u_1) g(u_2; i_1, k_1) a(t_{i_2} - u_1) g(u_2; i_2, k_2) du_1 du_2, \\ A_{14} &= \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \int_{u_1} \int_{u_2} a(t_{i_1} - u_1) g(u_2; i_1, k_1) a(t_{i_2} - u_2) g(u_1; i_2, k_2) du_1 du_2. \end{aligned}$$

Clearly,

$$A_{12} = A_2.$$

By the triangle inequality, (39), and condition (C1),

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

Similarly,

$$\begin{aligned}
|A_{13}| &\leq \sum_{i_1 \in I_{k_1}} \sum_{i_2 \in I_{k_2}} \int_{u_1} \int_{u_2} \beta(t_{i_1} - u_1)\beta(t_{i_1} + k_1\pi/\nu - u_2) \\
&\qquad \times \beta(t_{i_2} - u_1)\beta(t_{i_2} + k_2\pi/\nu - u_2) du_1 du_2 \\
&\leq C \sum_{i \in I_{k_1}} \sum_{i_2 \in I_{k_2}} B(t_{i_1} - t_{i_2})B(t_{i_1} - t_{i_2} + (k_1 - k_2)\pi/\nu) \\
&\leq C n_{k_1} \sum_i B(t_i)B(t_i + (k_1 - k_2)\pi/\nu),
\end{aligned}$$

and

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

Thus,

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

where

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

It follows from (19) and the assumption $\lambda \geq N^{-1}$, that

$$\frac{1}{n_k + k^2\lambda} \leq C \max\left(\frac{1}{N}, \frac{1}{N^2\lambda}\right) \leq \frac{C}{N}, \text{ for all } k. \tag{40}$$

Since

$$\frac{1}{\nu}\sum_k B(k\pi/\nu + k'\pi/\nu) = \frac{1}{\nu}\sum_k B(k\pi/\nu) < \infty, \text{ for all } k',$$

it follows from (40) that

$$\sum_{k=0}^{K} |b_k| B(k\pi/\nu + k'\pi/\nu) = \frac{1}{\nu}\sum_k \frac{1}{n_k + k^2\lambda} B(k\pi/\nu + k'\pi/\nu) \leq \frac{C}{N}. \tag{41}$$

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

$$T_1 \leq \frac{C}{\sqrt{N\lambda}}.$$

The same can be concluded for T_2 and T_3 using similar derivations, and so

$$\text{var}(\hat{f}_\lambda(\omega)) \leq \frac{C}{\sqrt{N\lambda}}.$$

This concludes the derivation for the bound of the variance.

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

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

where

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

First, by (23) and condition (C2),

$$U_1 \le C\int_{K/\nu}^{\infty}(u\pi)^{-\alpha-1}du \le C\left(\frac{\nu}{K}\right)^{\alpha}. \tag{42}$$

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

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

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

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

which is finite by condition (C3). Thus,

$$U_2 \le \frac{C}{\nu}. \tag{43}$$

Now consider U_3. By (23) and condition (C2),

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

where $[\nu]$ denotes the integer part of ν. By (40),

$$V_1 = \frac{C}{\nu}\sum_{k=1}^{[\nu]}\frac{k^2\lambda}{n_k+k^2\lambda}\le\frac{C\lambda\nu^2}{N}. \tag{44}$$

Also, by (19),

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

Observe that

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

Since $\nu \ge \nu_0$, it is easy to see that the two terms on the right 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

$$V_2 \le C \left[\frac{\lambda\nu^2}{N} + \left(\frac{\nu}{N}\right)^\alpha \right]. \tag{45}$$

Summarizing the results from (42)-(45), we conclude that

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

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