UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 07-26
Business Economics Series 06
May 2007, corrected in 2009

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249607

AUTOMATIC SPECTRAL DENSITY ESTIMATION FOR RANDOM FIELDS ON A LATTICE VIA BOOTSTRAP[*]

Jose M. Vidal-Sanz[1]

Abstract

This paper considers the nonparametric estimation of spectral densities for second order stationary random fields on a d-dimensional lattice. I discuss some drawbacks of standard methods, and propose modified estimator classes with improved bias convergence rate, emphasizing the use of kernel methods and the choice of an optimal smoothing number. I prove uniform consistency and study the uniform asymptotic distribution when the optimal smoothing number is estimated from the sampled data.

Keywords: Spatial data, spectral density, smoothing number, uniform asymptotic distribution, Bootstrap.
AMS 2000 subject classifications:. Primary 62M30; secondary 62M15, 62G20

[*] A version of this paper is published in TEST, by Springer Verlag, DOI 10.1007/s11749-007-0059-5. I wish to thank Professor C. Velasco and two anonymous referees for their helpful comments and suggestions, and Professor P. M. Robinson for introducing me in the topic. This research has been supported by a Marie Curie Grant, Mobility 11, of the European Commission, with reference number FP6-2004-505469..

[1] E-mail: jvidal@emp.uc3m.es

1. INTRODUCTION

The estimation of the power spectrum for random fields on a d-dimensional lattice is relevant for many purposes, including specification and testing of parametric models, detecting anisotropies and hidden periodicities, signal extraction from noisy random fields, interpolation, prediction and smoothing. It is also useful to obtain a more sparse decomposition of a digital image, requiring less storage space. Spatial spectral methods have been applied to ecological data (e.g., Reshaw, 1984, and Reshaw and Ford, 1983), earth sciences (Agterberg, 1967), astronomy (Abramenko et al, 2001), and meteorology (Barry and Perry, 1973), among others.

This paper is concerned with nonparametric estimation of the spectral density for spatial processes. I discuss some drawbacks in the current estimation methods. The bias of nonparametric estimators based on Whittle's (1954) periodogram achieve low convergence rate due to the "edge effects," whilst the smoothed periodogram based on Guyon's (1982) periodogram can present consistency problems. I overcome these problems by smoothing a modified periodogram introduced by Robinson and Vidal-Sanz (2006). I focus on kernel estimators, for which the choice of an optimal smoothing number is considered. Furthermore, the uniform consistency and uniform asymptotic distribution are established when the optimal smoothing number is estimated from the data (see Theorem 3). The uniform asymptotic distribution result has also interest in time series context, complementing Robinson's (1991) uniform consistency result for automatic estimation. Finally, I present a consistent Bootstrap method for the automatic estimation of the smoothing number.

Consider a real second-order stationary stochastic process on a d-dimensional lattice, $\left\{X_t : t \in \mathbb{Z}^d\right\}$, where $\mathbb{Z} = \{0, \pm 1, \pm 2, ..\}$, with first moments $E\left[X_t\right] = \mu$ and $\gamma_l = Cov\left[X_t X_{t+l}\right]$. I will assume that there exists an integrable spectral density $f\left(\lambda\right) \geq 0$ on $\Pi^d = \left[-\pi, \pi\right]^d$, such that $\gamma_l = \int_{\Pi^d} e^{il\cdot\lambda} f\left(\lambda\right) d\lambda$, with $l \cdot \lambda = \sum_{r=1}^d \lambda_r l_r$, and $f\left(\lambda\right) = \left(2\pi\right)^{-d} \sum_{l\in\mathbb{Z}^d} \gamma_l e^{-il\cdot\lambda}$ (this is why f is also called the power spectrum). The spectral density can be periodically extended to $\mathbb{R}^d$. A sufficient condition for the existence of f is that $\sum_{l\in\mathbb{Z}^d} |\gamma_l| < \infty$, this also implies that $f \in C\left(\Pi^d\right)$ and it obeys the Lipschitz condition $f \in Lip\left(\alpha\right)$ for any $\alpha < 1/2$, where $f \in Lip\left(\alpha\right)$ means that $\sup_{0<\|h\|\leq\delta} \left\|f\left(\lambda\right) - f\left(\lambda + h\right)\right\|_\infty = o\left(\delta^\alpha\right)$ when $\delta \downarrow 0$ with $\|f\|_\infty = \sup_{\lambda\in\Pi^d} \left|f\left(\lambda\right)\right|$. Under the stronger condition $\sum_{l\in\mathbb{Z}^d} \left(1 + \|l\|_1^q\right) |\gamma_l| < \infty$ for some integer $q \geq 1$, where $\|l\|_1 = \sum_{r=1}^d |l_r|$, we have that $f \in C^q\left(\Pi^d\right)$.

In spatial data, it is customary to take the beginning data situated at the origin (or at one), but sometimes data are centered elsewhere, and asymptotic could require that the sample increases in all directions of space. Therefore, without loss of generality, I consider the estimation of the spectral density when X_t is observed on a rectangular set

$$N = \left\{ t \in \mathbb{Z}^d : -n_r^L \leq t_r \leq n_r^U, \quad r = 1, ...d \right\},$$

where $n_r^L, n_r^U \in \mathbb{Z}$ and $-n_r^L \leq n_r^U$ for $r = 1, ...d$. Then, define $n_r = n_r^L + n_r^U + 1$ and $n = \prod_{r=1}^d n_r$ the cardinal of N. Following Robinson and Vidal-Sanz (2006), for the asymptotic regime we regard $n_r = n_r(n)$ as a function of the total number of observations n, which is the basic index for asymptotic results; and we require that n_r increases for all r introducing the following assumption,

A.1. *For all sufficiently large n, there exist $\xi > 0$, $c > 0$ such that*

$$n_r(n) > c_1 n^{\xi}.$$

Since $\sum_{r=1}^d n_r(n)^{-1} \geq d\left(\prod_{r=1}^d n_r^{-1}\right)^{1/d} = dn^{-1/d}$ we have that $\xi \leq 1/d$ where the equality[1] is attained when all $n_r(n)$ increase at the same rate $n^{1/d}$. This specification covers many possibilities. For example, we could set $n_r^L = -1$ for $r = 1, .., d$ and therefore consider the standard unilateral sample case $N = \times_{r=1}^d \left\{1, ..., n_r^U\right\}$ with $n = \prod_{r=1}^d n_r^U$, and $n_r^U \to \infty$. The spatial statistics literature focuses on this case, but spatial samples could generally increase in one or several directions. For example, we can observe a symmetric sample with $n_r^L = -n_r^U$, $n_r^U \geq 0$ for all coordinates, so that $N = \times_{r=1}^d \left\{-n_r^U, ..., n_r^U\right\}$ with $n = \prod_{r=1}^d \left(2n_r^U + 1\right)$.

For any $l \in \mathbb{Z}^d$ let us define $N(l) := \{t \in N : t + l \in N\}$ with cardinal $n(l) := \prod_{r=1}^d (n_r - |l_r|)$, the unbiased covariance estimator

$$c_{n,l}^* = \frac{1}{n(l)} \sum_{t \in N(l)} \left(X_t - \widehat{\mu}_n\right)\left(X_{t+l} - \widehat{\mu}_n\right),$$

with $\widehat{\mu}_n = n^{-1} \sum_{t \in N} X_t$, the biased covariance estimator $c_{n,l} = w(l)\, c_{n,l}^*$ with $w(l) = n(l)\, n^{-1} = \prod_{r=1}^d (1 - |l_r|/n_r)$, and the discrete Fourier transform $d_n(\lambda) = \left(n(2\pi)^d\right)^{-1/2} \sum_{t \in N} \left(X_t - \widehat{\mu}_n\right) e^{-it \cdot \lambda}$.

Whittle (1954) introduced the spatial periodogram in the context of unilateral samples. The spatial periodogram,

$$I(\lambda) = |d_n(\lambda)|^2 = (2\pi)^{-d} \sum_l{}' \, c_{n,l}\, e^{-il \cdot \lambda} = (2\pi)^{-d} \sum_l{}' \, w(l)\, c_{n,l}^*\, e^{-il \cdot \lambda},$$

where $\sum_l{}'$ denotes the sum for $l \in \mathbb{Z}^d$ such that $|l_r| \leq n_r - 1$, $r = 1, ..., d$, is asymptotically unbiased for $f(\lambda)$. But the variance of $I(\lambda)$ does not tend to zero, as it can be anticipated, and some

[1] Warning: There is a typo in the published version of the paper, where the inequality $\sum_{r=1}^d n_r(n)^{-1} \geq d\left(\prod_{r=1}^d n_r^{-1}\right)^{1/d}$ is written in reverse sense.

smoothing is required. Henceforth, I will use the discrete frequencies $\omega_{j,n} = (2\pi j_1/n_1, ..., 2\pi j_d/n_d)$, for all $j \in J_n$, where the set $J_n = \times_{r=1}^{d} \{0, ..., n_r - 1\}$ has cardinal n. The numerical effort required to compute $I(\omega_{j,n})$ can be reduced by using the planar Fast Fourier Transform, see Reshaw and Ford (1983) for a discussion.

Spatial literature has discussed the nonparametric spectral density estimation for random fields with samples spreading in one direction ($n_r^L \equiv -1$, $n_r^U \to \infty$), see e.g. Priestley (1964), Rozanov (1967), Brillinger (1970), Rosenblatt (1985), Ivanov and Leonenko (1986), Žurbenko (1986), Heyde and Gay (1993), and Leonenko (1999), among others. The basic theory is a straightforward generalization from time series. One of the most simple estimators is the class of *smoothed periodogram* estimators,

$$\widehat{f}(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I(\omega_{j,n}),$$

where M is a smoothing number. The weight functions $\{K_M\} \subset L_1(\mathbb{R}^d)$ are symmetric, continuous and periodical with periodicity $[0, 2\pi]^d$, and, as the smoothing parameter $M \to \infty$ the functions $K_M \to \delta_0$ (where δ_0 is the periodic extension of the Dirac's delta generalized function at 0). Consistency requires M depending on n, with $M_n \to \infty$ at a rate sufficiently slow to ensure that the variance of $\widehat{f}$ tends to zero. Another popular class of spectral density estimators known as *lag windowed* estimators, is defined by

$$\widetilde{f}(\lambda) = (2\pi)^{-d} \sum_{l}{}' \, k_M(l) \, w(l) \, c_{n,l}^{*} e^{-il\cdot\lambda}, \tag{1}$$

where $k_M(l)$ is the lag window, satisfying $k_M(l) = k_M(-l) \leq k_M(0) = 1$ and the parameter M plays the role of a smoothing number. It is possible to consider different kinds of smoothing numbers. When $M \in \mathbb{N}^d$ and $k_M(l) = 0$ for $|l_r| \geq M_r$ and $r = 1, ..., d$ the parameters M are called lag numbers. The smoothing numbers could be positive definite matrices $M \in \mathbb{R}^{d \times d}$ such that $k_M(l) = k(M^{-1}l)$ with $|k(l)| \leq k(0) = 1$ for all l and $k(l) = 0$ for $|l| \geq 1$. For diagonal matrices the vector $diag(M)$ can be regarded as lag numbers. Lag windowed and smoothed periodogram estimators can be related. Introducing $K_M(u) = (2\pi)^{-d} \sum_l{}' \, k_M(l) \, e^{-il\cdot u}$, we can express lag windowed estimators as

$$\widetilde{f}(\lambda) = \int_{\Pi^d} \left((2\pi)^{-d} \sum_{l}{}' \, k_M(l) \, e^{-il\cdot(\lambda - u)} \right) I(u) \, du = \int_{\Pi^d} K_M(u - \lambda) I(u) \, du,$$

where $k_M(l) = \int_{\Pi^d} e^{il\cdot\lambda} K_M(\lambda) \, d\lambda$. Thus, $\widehat{f}$ can be thought of as a numerical integration approximation to $\widetilde{f}$.

For any of these estimator classes, the consistency can be established much as in the time series literature. Unfortunately, the spatial density estimators previously discussed are exposed to a low bias convergence rate, inherent in the Whittle spatial periodogram. As $E[I(\lambda)]$ is the Cesaro sum of the multiple Fourier series of f, (see e.g., Zygmund, 1959, Vol. II, Chapter XVII),

$$E[I(\lambda)] = (2\pi)^{-d} \sum_{l}{}' \, w(l)\,\gamma_l e^{-il\cdot\lambda} = \int_{\Pi^d} F_n(u-\lambda)\, f(u)\, du,$$

where $F_n(u) = \prod_{r=1}^{d} (2\pi n_i)^{-1} \left(\sin\{n_r u_r/2\} / \sin\{u_r/2\}\right)^2$, is the multivariate Fejer kernel. Let us consider $\omega(f,\delta) = \sup_{0<\|h\|\leq\delta} \|f(\lambda) - f(\lambda+h)\|_\infty$. As a consequence of the Korovkin Theorem (see, e.g., Korovkin, 1960), we have that, as $n \to \infty$,

$$\|E[I(\lambda)] - f\|_\infty \leq 2\omega(f,\delta_n) = o\left(\delta_n^{1/2}\right)$$

for all $f \in Lip(\alpha)$ with $\alpha > 1/2$, where

$$\begin{aligned} \delta_n &= \int_{\Pi^d} F_n(u-\lambda)\, \|u\|\, du \leq K \prod_{i=1}^{d} \int_{-\pi}^{\pi} \frac{1}{2\pi n_i} \left(\frac{\sin\{n_i u_i/2\}}{\sin\{u_i/2\}}\right)^2 \left(\sum_{r=1}^{d} |u_r|\right) du \\ &= K \sum_{r=1}^{d} \int_{-\pi}^{\pi} \frac{1}{2\pi n_i} \left(\frac{\sin\{n_r u_r/2\}}{\sin\{u_r/2\}}\right)^2 |u_r|\, du = O\left(\sum_{i=1}^{d} n_r^{-1}\right), \end{aligned}$$

which by Assumption $A.1.$ is of order not less than $n^{-1/d}$, and the uniform bias rate of I can be lower than $o(1/\sqrt{n})$ for $d > 1$. The basic reason for the low convergence rate is the edge effect, noticed by Guyon (1982). For a fixed l, as all $n_r \to \infty$ the bias $|E[c_{n,l}] - \gamma_l|$ is of order $\sum_{r=1}^{d} n_r^{-1} \geq dn^{-1/d}$. Thus, for a continuous integrable kernel K,

$$\left| \frac{1}{n} \sum_{j\in J_n} K(\omega_{j,n} - \lambda)\left(E[I(\omega_{j,n})] - f(\omega_{j,n})\right) \right| = o\left(n^{-\xi/2}\right),$$

by Assumption $A.1$. Therefore, the uniform rate of convergence is $o(1/\sqrt{n})$ only for $d = 1$ but can be significantly slower for $d > 1$.

To avoid the edge effect, Guyon (1982) introduced the modified periodogram with unbiased covariances,

$$I_*(\lambda) = (2\pi)^{-d} \sum_{l}{}' \, c_l^* e^{-il\cdot\lambda},$$

for unilateral samples. Note that $I_*(\lambda)$ is not necessarily a nonnegative function, and $E[I_*]$ is the multiple Fourier series of f. Although, there are infinitely many continuous functions f which Fourier series diverges to infinite (see e.g., Rudin, 1974, and Vidal-Sanz, 2005), $\|E[I_*(\lambda)] - f(\lambda)\|_\infty \to 0$ if f is a continuous function with bounded variation on Π^d. The modified periodogram I_* can be smoothed to estimate the spectral density f when it is enough regular. Politis and Romano

(1996) suggested to use unbiased autocovariances in spectral density estimation. The *lag windowed* estimator based on I_* is

$$\widetilde{f}_*(\lambda) = \int_{\Pi^d} K_M(u-\lambda) I_*(u)\, du = (2\pi)^{-d} \sum_l{}' k_M(l)\, c^*_{n,l} e^{-il\cdot\lambda},$$

similar to (1) with lag window $\{k_M(l)/w(l)\}$.

The theoretical properties of I_* have been criticized by Robinson and Vidal-Sanz (2006), in the context of Whittle estimation, due to the presence of aliasing problems. This problem can also be found in *smoothed periodogram* estimators; it suffices to consider the weight function $K_M(\lambda) = (2\pi)^{-d} \sum_l k_M(l)\, e^{-il\cdot\lambda}$. Applying Hannan´s (1973) argument, we have that

$$\widehat{f}_*(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_*(\omega_{j,n}) = \sum_l{}' k_M(l) \left(c^*_l + c^*_{l\pm n}\right), \tag{2}$$

where $c^*_{l\pm n} = 0$ for $l = 0$ and $c^*_{l\pm n} = c^*_{n-l}$ for $l \neq 0$. The right-hand side of (2) is equal to

$$= (2\pi)^{-d} \int_{\Pi^d} K_M(\omega - \lambda) I_*(\omega)\, d\omega + \sum_l{}' k_M(l)\, c^*_{l\pm n},$$

where $c^*_{l\pm n}$ is composed of at most $n-l$ terms of the form $X_t X_{n-l+t}$ divided by l, which does not converge to zero (e.g., $c^*_{n-1} = X_1 X_n$). Although $k_M(l) \to 0$, if this convergence is not uniform in l an smoothed periodogram based on I_* could be inconsistent or, in the best case, the rate of convergence could be too slow. By contrast, in the Whittle periodogram $c_{l\pm n} = O_p\left(n^{-1}\right)$ and the "aliasing" of lags does not generate the inconsistency, as proved by Hannan (1973).

Dahlhaus and Künsch (1987) proposed to use a periodogram I_T the covariances of which use tapered data, using this periodogram for Whittle estimation of parametric models. They show that, for $d \leq 3$, if the taper uses an appropriate bandwidth the estimated parameters are consistent with rate $\sqrt{n}$. Robinson (2007) suggested to use tappered periodograms in spectral density estimation. But for lag windowed spectral estimators based on a such periodogram it would be required to choose a taper, a bandwidth, and a smoothing number; introducing too much ambiguity in the estimation.

In this paper a modified spectral density estimator is presented which is not affected by the aliasing, nor the edge effect. In Section 2, the modified estimators are introduced focusing on kernel estimators. Also, the optimal smoothing number are considered for the integrated mean-square loss function, which is infeasible and has to be estimated from the sample data. The issue of spatial sampling interval also is discussed. Section 3 contains the main theoretical results. For a stochastic smoothing number the uniform consistency and pointwise asymptotic normality of modified kernel estimators are proved. In Section 4 consistency of plug-in and Bootstrap estimators of the optimal smoothing number is considered. Proofs are included in the Appendix.

2. MODIFIED SPECTRAL DENSITY ESTIMATORS

To avoid the aliasing problems in I_*, Robinson and Vidal-Sanz (2006) introduced a truncated periodogram,

$$I_g(\lambda) = (2\pi)^{-d} \sum_{\substack{l \in \mathbb{Z}^d : |l_r| \leq g(n_r) \\ r=1,\ldots d}} c_l^* e^{-il\cdot\lambda},$$

where g is a function satisfying:

A.2. g is a positive, integer valued, monotonically increasing function such that $g(x) \to \infty$ as $x \to \infty$, and for for some $c_2 \in (0,1)$ $g(x) \leq c_2 x$ for all $x > 0$.

For example we can take $g(x) = [\alpha x]$ with $\alpha \in (0,1)$ and $[\cdot]$ the integer part; which in practice means that we consider a trimmed summation of elements l with coordinates $|l_r| \leq \alpha n_r$. The advantage of this approach is that the parameters in function g do not play an asymptotic effect, by contrast with tapering methods. Some finite sample experiments can be found in Robinson and Vidal-Sanz (2006).

Robinson and Vidal-Sanz (2006) prove that $\|E[I_g] - f\| = o\left(n^{-1/2}\right)$ under appropriate assumptions on the covariance function (A.3. and A.4. in Section 3); and when averaged over discrete frequencies, the modified periodogram I_g is immune to aliasing problems affecting I_*. Here, it is proposed the class of *modified smoothed periodogram* estimators,

$$\widehat{f}^g(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_g(\omega_{j,n}), \tag{3}$$

and the class of *modified lag windowed* estimators

$$\widetilde{f}^g(\lambda) = \int_{\Pi^d} K_M(u - \lambda) I_g(u)\, du = (2\pi)^{-d} \sum_{|l_r| \leq g(n_r)} \cdots \sum_{r=1,..,d} k_M(l)\, c_l^* e^{-il\cdot\lambda} \tag{4}$$

with $K_M(u) = (2\pi)^{-d} \sum_{r=1}^d \sum_{|l_r| \leq g(n_r)} k_M(l)\, e^{-il\cdot u}$, and $k_M(l) = \int_{\Pi^d} e^{il\cdot\lambda} K_M(\lambda)\, d\lambda$. Both estimators $\widehat{f}^g$ and $\widetilde{f}^g$ are similar to the classical ones, but using I_g (instead of I or I_*) it is possible to establish the uniform consistency and derive the uniform weak distribution under appropriate conditions. As I_*, also I_g can take negative values for some frequencies, and so do $\widehat{f}^g$ and $\widetilde{f}^g$. Although negative frequency estimations are unlikely for large samples, we can vanish the estimator for frequencies with negative estimated power spectra by taking $\widehat{f}^{g+}(\lambda) = \max\left\{0, \widehat{f}^g(\lambda)\right\}$, i.e., the L_1-projection of $\widehat{f}^g$ onto the positive cone. $\widetilde{f}^{g+}$ is defined analogously.

A rigorous treatment of the asymptotic theory is given in Section 3, but some heuristic arguments are presented in this section. Proceeding much as in the time-series literature, under appropriate conditions the estimator $\widetilde{f}^g$ roughly satisfies

$$\begin{aligned} E\left[\widetilde{f}^g(\lambda)\right] &= \int_{\Pi^d} K_M(u-\lambda) f(u)\,du + o\left(n^{-1/2}\right), \\ Cov\left[\widetilde{f}^g(\lambda), \widetilde{f}^g(\theta)\right] &\approx \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u-\lambda) K_M(u-\theta) f(u)^2\,du, \end{aligned}$$

and $\widehat{f}^g$ exhibits an analogous behaviour, as the aliasing of lags does not affect the modified smoothed periodogram.

Applying the Korovkin Theorem, it can be proved that $\widetilde{f}^g$ is asymptotically unbiased for integrable and continuous f, and a Lipschitz assumption can be used to obtain a convergence rate. If $f \in C^r\left(\Pi^d\right)$, taking the Taylor expansion of $f(u+\lambda)$ in λ we obtain that

$$\begin{aligned} & E\left[\widetilde{f}^g(\lambda)\right] - f(\lambda) = \int_{\Pi^d} K_M(u)\left(f(u+\lambda) - f(\lambda)\right) du \\ = & \sum_{j=1}^{r-1} \sum_{\|\nu\|_1 = j} \frac{D^\nu f(\lambda)}{\nu!} \int_{\Pi^d} K_M(u) u^\nu du \\ & + r \sum_{\|\nu\|_1 = r} \frac{1}{\nu!} \int_{\Pi^d} \int_0^1 (1-t)^{r-1} D^\nu f(\lambda + tu) K_M(u) u^\nu dt du. \end{aligned}$$

we say that the family $\{K_M(u)\}$ is of order r if, for all M, we have that $\int_{\Pi^d} K_M(u) u^\nu du = 0$ for $1 \leq \|\nu\|_1 < r$ and $\int_{\Pi^d} \|u\|^r |K_M(u)|\,du < \infty$; this implies that the bias convergence rate to zero equals the rate of the remaining term, namely, $O\left(\int_{\Pi^d} \|u\|^r |K_M(u)|\,du\right)$, uniformly in frequency. In particular, the symmetry $k_M(l) = k_M(-l)$ implies that $\int_{\Pi^d} u K_M(u)\,du = 0$, and the bias rate is $O\left(\int_{\Pi^d} \|u\|^2 |K_M(u)|\,du\right)$ for $f \in C^2\left(\Pi^d\right)$. In some particular cases (e.g., kernel estimators) it is easy to obtain orders higher than 2, but it is not for general estimators. Delgado and Vidal-Sanz (2001) present a general methodology for obtaining families $\{K_M(\cdot)\}$ with higher orders.

Regarding the covariance structure, if K_M is supported on a closed neighborhood around the origin, the covariance tends to zero for $\lambda \neq \theta$, and the variance satisfies

$$\begin{aligned} Var\left[\widetilde{f}^g(\lambda)\right] &\approx \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u)^2 f(u+\lambda)^2\,du \\ &\approx f(\lambda)^2 \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u)^2\,du. \end{aligned}$$

This approximation is accurate for M, n be large, or f flat around λ. When $\int_{\Pi^d} K_{M_n}(u)^2\,du = o(n)$ the estimator will be mean-square consistent.

Several functional norms $\|\cdot\|$ can be used to study the global convergence $\left\|\widehat{f^g}-f\right\| \rightarrow 0$ in probability, i.e. different function spaces can be considered. Perhaps the most popular choices are $C\left(\Pi^d\right)$ endowed with the supremum norm $\|f\|_{\infty}=\sup_{\lambda\in\Pi^d}|f(\lambda)|$, and the space $L_2(\mu)$ for some Borel measure μ on Π^d, endowed with the mean square norm $\|f\|_{L_2(\mu)}=\left(\int_{\Pi^d}|f(\lambda)|^2\mu(d\lambda)\right)^{1/2}$, where the Lebesgue measure is frequently taken. Both are complete and separable Banach spaces, and $C\left(\Pi^d\right)$ is dense in $L_2(\mu)$. The uniform consistency is stronger than the L_2 consistency on Π^d and it will be considered in Section 3.

2.1. Kernel estimators

Perhaps the most relevant methods are *(modified) kernel estimators*, and the rest of the paper is focused on them. There are two alternative approaches to introduce kernel estimators. In the first one, kernel estimators are a class of smoothed periodograms (3), whilst in the second one, they are lag windowed methods (4). The distinctive aspect of kernel methods is that the kernel K_M is defined by

$$K_M(u)=\det(M)\sum_{l\in\mathbb{Z}^d}K(M(u+2\pi l)), \qquad u\in\Pi^d,$$

for a kernel function $K\in L_1\left(\mathbb{R}^d\right)$ with $\int_{\mathbb{R}^d}K(u)\,du=1$. The kernel K can be defined as the product of univariate kernels, $K(u)=\prod_{r=1}^d K_r(u_r)$, with $\{K_r\}\subset L_1(\mathbb{R})$. The smoothing number M_n is a sequence of symmetric positive definite matrices, with $M_n\rightarrow\infty$ and $\det(M_n)/n\rightarrow 0$. The kernel lag window $k_M(l)$ verifies,

$$\begin{aligned}
k_M(l) &= \int_{\Pi^d}e^{il\cdot u}K_M(u)\,du=\det(M)\sum_{l\in\mathbb{Z}^d}\int_{\Pi^d}e^{il\cdot u}K(M(u+2\pi l))\,du \\
&= \det(M)\int_{\mathbb{R}^d}e^{il\cdot u}K(Mu)\,du=\int_{\mathbb{R}^d}e^{il\cdot M^{-1}u}K(u)\,du=k\left(M^{-1}l\right),
\end{aligned}$$

with $k(x)=\int_{\mathbb{R}^d}e^{ix\cdot u}K(u)\,du$. Therefore, if $k\in L_2\left(\mathbb{R}^d\right)$, applying Parseval's equality, we have

$$\int_{\Pi^d}K_M(u)^2\,du=\frac{\det(M)}{(2\pi)^d}\int_{\mathbb{R}^d}k(u)^2\,du,$$

and $\int_{\Pi^d}K_{M_n}(u)^2\,du=o(n)$ as $\det(M_n)/n\rightarrow 0$.

Let us consider the matrix norm $\|M\|=(megv(M'M))^{1/2}$, where $megv$ means the maximum eigenvalue. We say that K is a kernel of order q if $\int_{\Pi^d}K(u)\,u^{\nu}du=0$ for $1\leq\|\nu\|_1<q$ and $\int_{\Pi^d}|K(u)|\,\|u\|^q\,du<\infty$. The q-order property ensures that for $f\in C^q\left(\Pi^d\right)$ the spectral density bias is $O\left(\|M_n\|^{-q}\right)$ uniformly in frequency. This high order rate is a relevant property in order to ensure that the bias tends to zero at rate $o\left(n^{-1/2}\right)$. Note that K is of order $q = 2$ whenever it is even and $\int_{\Pi^d} |K(u)| \|u\|^2 du < \infty$. For K to be of order $q > 2$, it is necessary that K takes negative values. The q-order kernel property can be stated by the requirement that: $k(x)$ is q-times continuously differentiable at zero with $D^{\nu} k(x)|_{x=0} = 0$ for all integer vectors $0 < \|\nu\|_1 < q$. Since $k(x) \leq k(0) = 1$, taking into account the Taylor expansion definition the last condition can be equivalently expressed by the following condition:

$$\lim_{x \to 0} \frac{1 - k(x)}{\|x\|^q} = k_q,$$

for some finite constant k_q. The extreme case $q = \infty$, is often identified with the "flat-top" kernels considered by Politis and Romano (1996).

The variance of kernel estimators is $O(\det(M_n)/n)$ and the square bias is at best $O\left(\|M_n\|^{-2q}\right)$, both rates are satisfied uniformly in frequency. If $M_n = m_n S$ with m_n scalar and S a symmetric positive definite matrix, then the mean-square error of kernel estimation is $O\left(m_n^{-2q} + m_n^d/n\right)$ uniformly in frequency, and the rate of convergence is made as fast as possible by taking $m_n = O\left(n^{1/(2q+d)}\right)$, with associated mean-square error $O\left(n^{\frac{-2q}{2q+d}}\right)$. In particular, when $q = d = 2$, the optimal rate is $m_n = O\left(n^{1/6}\right)$, which suggests that we could take $M_n = Sn^{1/6}$ for a matrix S, and the associated mean square error is $O\left(n^{-4/6}\right)$. The curse of dimensionality can be observed, as the mean square error rate $n^{2q/(2q+d)}$ decreases exponentially when the dimension d increases, implying that for high dimensions the sample size n required for accurate estimations should be increasingly large. In space-temporal context we rarely find dimensions $d > 4$, and actual sample sets are usually large enough to avoid concerns about this issue.

2.2. Kernel Optimal Smoothing numbers

The choice of the parameter S is crucial to deal with the trade-off effects between the bias and variance in finite samples, and it should be based on some objective loss function. Different loss functions lead to different optimal parameters S^*, that usually depend on the unknown f, but there is not a universally optimal parameter for all loss criteria. A relevant and manageable loss function is the integrated mean-square error with respect to the weight measure μ,

$$\begin{aligned} IMSE(M, n, \mu) &= E\left[\left\|\widetilde{f}^g - f\right\|^2_{L_2(\mu)}\right] = \int_{\Pi^d} E\left[\left|\widetilde{f}^g(\lambda) - f(\lambda)\right|^2\right] \mu(d\lambda) \\ &= \int_{\Pi^d} E\left[\left|\widetilde{f}^g(\lambda) - E\left[\widetilde{f}^g(\lambda)\right]\right|^2\right] \mu(d\lambda) + \int_{\Pi^d} \left|E\left[\widetilde{f}^g(\lambda)\right] - f(\lambda)\right|^2 \mu(d\lambda), \end{aligned}$$

by Fubini´s theorem. If $f \in C^2\left(\Pi^d\right)$ and K is of order 2, the bias is

$$
\begin{aligned}
Bi\left[\widetilde{f}^g(\lambda)\right] &= \int_{\Pi^d} K(u)\left(f(\lambda + M_n u) - f(\lambda)\right) \\
&= \int_{\Pi^d} K(u)\left(u' M_n \nabla f(\lambda)\, du + \frac{1}{2} u' M_n^{-1\prime} \frac{\partial^2}{\partial\lambda\partial\lambda'} f(\lambda) M_n^{-1} u\right) du + o\left(\|M_n\|^{-2}\right) \\
&= \frac{1}{2} Tr\left\{\int_{\Pi^d} u\ u' K(u)\, du\ \cdot M_n^{-1\prime} \frac{\partial^2}{\partial\lambda\partial\lambda'} f(\lambda) M_n^{-1}\right\}^2 \mu(d\lambda) + o\left(\|M_n\|^{-2}\right),
\end{aligned}
$$

where the $o\left(\|M_n\|^{-2}\right)$ term is uniform in λ, and Tr denotes the trace of a square matrix.

Proceeding heuristically (a precise treatment is presented in Section 3), we have that

$$
Var\left[\widetilde{f}^g(\lambda)\right] \approx \frac{\det(M_n)}{n} \kappa^2 f(\lambda)^2,
$$

with $\kappa^2 = \int_{\mathbb{R}^d} k(u)^2\, du$, and the $o(\cdot)$ term is uniform in λ, and taking $M_n = m_n S$,

$$
IMSE(m_n S, n, \mu) \approx \frac{m_n^d \det(S)}{n} \kappa^2 F^2 + \frac{C_K^2}{4 m_n^4} \int_{\Pi^d} \left| Tr\left\{ (SS')^{-1} \frac{\partial^2}{\partial\lambda\partial\lambda'} f(\lambda) \right\} \right|^2 \mu(d\lambda)
$$

where $C_K = \int_{\Pi^d} u\ u' K(u)\, du$, and $F^2 = \int_{\Pi^d} f(\lambda)^2 \mu(d\lambda)$. If we use the optimal rate for $q = 2$, i.e. $m_n = n^{1/(4+d)}$, then for n large

$$
n^{4/(4+d)} IMSE\left(S n^{1/(4+d)}, n, \mu\right) \approx \det(S) \kappa^2 F^2 + \frac{C_K^2}{4} \int_{\Pi^d} \left| Tr\left\{ (SS')^{-1} \frac{\partial^2}{\partial\lambda\partial\lambda'} f(\lambda) \right\} \right|^2 \mu(d\lambda).
$$

The right hand side can be minimized in S, taking

$$
S_0^* = \left(\frac{4 C_K^2}{\kappa^2 F^2} \int_{\Pi^d} \left(\frac{\partial^2}{\partial\lambda\partial\lambda'} f(\lambda) \right)^2 \mu(d\lambda) \right)^{1/4+d}.
$$

Therefore, we do not use the same bandwidth in each dimension of the frequency space, but rather a general elliptically shaped kernel at a particular rotation controlled by $(S_0^* S_0^{*\prime})$. Analogous arguments can be applied for $E\left[\left\|\widehat{f}^g - f\right\|_{L_2(\mu)}^2\right]$. Higher order kernels can be considered, but f should satisfy higher differentiability requirements. In all the cases, the optimal value is a function, $S_0^* = S^*(f)$, of the unknown f.

Though $S_0^* = S^*(f)$ is infeasible, usually it can be estimated from the data by a plug-in procedure, some cross-validation method, or Bootstrap. The plug-in procedure takes a consistent pilot estimation $\widetilde{f}_{M_0}^g$, and estimates $\widehat{S}^* = S^*\left(\widetilde{f}_{M_0}^g\right)$. For example, when $f \in C^3\left(\Pi^d\right)$ and some regularity conditions are satisfied we can use a kernel pilot, as $\partial^2 \widetilde{f}^g / \partial\lambda\partial\lambda'$ is consistent respect to $\partial^2 f / \partial\lambda\partial\lambda'$. The plug-in procedure can be iterated. Cross validation methods are popular in time series analysis, see Beltrao and Bloomfield (1987) and Robinson (1991, Sec. 5), and they can be extended to deal with spatial data. However, in this paper I will focus on Bootstrap methods. Our approach is different from the Frank and Härdle (1992) time series bootstrap method for kernel spectral estimators, based on a Studentized periodogram. See Section 4 for details.

Summarizing, nonparametric estimation of power spectrum requires the choice of an appropriate smoothing number M_n. The choice of an optimal smoothing number entails the choice of a loss function leading to some optimal smoothing number, usually infeasible though it can be estimated from the sampled data. As a consequence, the smoothing number M_n should be allowed to depend on the data, provided that $\det(M_n)/n \to_p 0$ and $M_n \to_p \infty$, as required for mean-square consistency.

2.3. Sampling effects

Earth sciences often collect data from a continuous phenomena at regular intervals, using fixed monitoring points. Consider a real second-order stationary stochastic process $\left\{X_t : t \in \mathbb{R}^d\right\}$, with continuous spatial index, with spectral density $f \in C\left(\mathbb{R}^d\right)$, and covariances $\gamma_l = \int_{\mathbb{R}^d} f(\lambda) e^{il\cdot\lambda} d\lambda$. Assume that the sampling interval for each coordinate is $\Delta = (\Delta_1, ..., \Delta_d)^T$. For any $t \in \mathbb{Z}^d$ define $t \otimes \Delta = (t_1\Delta_1, ..., t_d\Delta_d)^T$ and $\Pi^d_\Delta = \prod_{r=1}^d [-\pi/\Delta_r, \pi/\Delta_r]$. The upper limit of the interval, $\left(\pi\Delta_1^{-1}, ..., \pi\Delta_d^{-1}\right)$, is known as the Nyqvist or folding frequency. Then, the sampled process, $\left\{X_{t\otimes\Delta} : t \in \mathbb{Z}^d\right\}$, has a spectral density f_Δ given by the folding formula,

$$f_\Delta(\lambda) = \sum_{j\in\mathbb{Z}^d} f(\lambda + \omega_{j,\Delta}),$$

where $\omega_{j,\Delta} = (2\pi j_1/\Delta_1, ..., 2\pi j_d/\Delta_d)$ are called alias and $\lambda \in \Pi^d_\Delta$. A peak on the spectrum f_Δ observed at frequency λ can be caused by an aliased frequency $\omega_{j,\Delta}$, unless f possesses no components with frequency greater than the Nyqvist frequency, i.e. $f_\Delta(\lambda) = f(\lambda)$ for $\lambda \in \Pi^d_\Delta$. Note that $\gamma_{l\otimes\Delta} = \int_{\Pi^d_\Delta} f_\Delta(\lambda) e^{i(l\otimes\Delta)\cdot\lambda} d\lambda$ for all $l \in \mathbb{Z}^d$ and $f_\Delta(\lambda) = \prod_{r=1}^d (\Delta_r/2\pi) \sum_{l\in\mathbb{Z}^d} \gamma_{l\otimes\Delta} e^{-i(l\otimes\Delta)\cdot\lambda}$. Using the observed data $\{X_{t\otimes\Delta} : t \in N\}$, a modified nonparametric estimator of f_Δ can be defined similarly to the case of unit sampling distance, i.e., smoothing the modified periodogram

$$I_{\Delta g}(\lambda) = \prod_{r=1}^d (\Delta_r/2\pi) \sum_{\substack{l\in\mathbb{Z}^d: |l_r| \le g(n_r)\\ r=1,...d}} c^*_{l\otimes\Delta} e^{-i(l\otimes\Delta)\cdot\lambda}.$$

The presented approach is valid to study the statistical behavior of the process on the regular sampling net, but something can be inferred about the continuous process when data are densely collected. Since $f(\lambda) \to 0$ as $\|\lambda\| \to \infty$ for an integrable f, for a sufficiently small Δ there are no appreciable components in f with frequencies higher than the Nyqvist frequency and the estimator

$\widehat{f}_{\Delta}$ can be used to infer approximately the behavior of f. The error decreases slowly only when f has heavy tails, i.e., when γ_l presents nonsmooth features.

3. MAIN RESULTS

This section is devoted to the uniform consistency and uniform asymptotic distribution of modified kernel spectral density estimators with multilateral samples. To derive the asymptotic theory I will assume a linear representation, introducing the following assumption,

A.3. The spatial process $\{X_t\}_{t\in\mathbb{Z}^d}$ follows a second order stationary random field with linear representation

$$X_t = \mu + \sum_{j\in\mathbb{Z}^d} \beta_j \, \varepsilon_{t-j},$$

where $\sum_{j\in\mathbb{Z}^d} \left|\beta_j\right| < \infty$, and $\{\varepsilon_j\}$ are identically and independently distributed random variables with zero mean, σ_ε^2 variance and forth order cumulant $\kappa_\varepsilon < \infty$.

Other approaches have been pursued in the literature. For example, we can assume conditions on the existence stationarity and summability of higher-order cumulants of $\{X_t\}$, using arguments related to Brillinger (1981). But for the estimation of second order spectra it is not really necessary to involve conditions on higher moments. Markovian assumptions or m-dependence conditions can be also considered to derive asymptotic results, but spatial correlations often decay slowly (see e.g. Ripley, 1988, p. 3). Mixing conditions are often used, see Doukhan (1994) for a review. Perhaps, Bolthausen's (1982) central limit theorem for α-mixing random fields is the most popular method. Linear processes as described in $A.3.$ are often used to justify the α-mixing assumption for X_t, under the requirement that the probability density function of ε_t satisfies a Lipschitz condition. I avoid this requirement, by following a martingale difference approach based on $A.3.$ I also assume that:

A.4. The spatial process $\{X_t\}_{t\in\mathbb{Z}^d}$ follows a second order stationary random field, which autocovariance function $\gamma_l = Cov\left[X_0, X_l\right]$ satisfies

$$\sum_{l\in\mathbb{Z}^d} \left(\sum_{r=1}^{d} g^{-1}\left(|l_r|\right)^{1/2\xi} \right) |\gamma_l| < \infty.$$

for ξ as in $A.1.$, and g^{-1} is the inverse function of g given in $A.2.$

A.5. K, k are continuous, real, even, integrable functions, and $\int K(u)\, du = 1$.

A.6. The lag window satisfies $\int \prod_{r=1}^{d} |u_r| \, |k(u)| \, du < \infty$.

A.7. The lag window satisfies $k(u) = 0$ *when some* $|u_r| > 1$, $r = 1, .., d$.

A.8. For some $q > 1$,

$$\lim_{x \to 0} \frac{1 - k(x)}{\|x\|^q} = k_q,$$

for some finite constant k_q.

Recall that if $M_n(S^0) = m_n S^0$, with m_n scalar and S^0 a symmetric positive definite matrix, then the mean-square error of kernel estimation is $O\left(m^{-2q} + m^d/n\right)$ uniformly in frequency, and the optimal rate of convergence is achieved by $m_n = n^{1/(2q+d)}$. Usually an optimal S is specified by some loss function, and consistently estimated. For a stochastic matrix $\widehat{M}_n = M_n\left(\widehat{S}_n\right) = m_n \widehat{S}_n$, where $\widehat{S}_n \to_p S^0$ and m_n is deterministic, I prove the uniform consistency of kernel estimators $\widehat{f}^g$ and $\widetilde{f}^g$ based on $\widehat{M}_n$.

Theorem 1 *Assume A.1. to A.5. and that* $\int |K(u)| \, \|u\| \, du < \infty$. *Consider* $\widehat{M}_n = m_n \widehat{S}_n$ *where* m_n *is a deterministic sequence and* $\widehat{S}_n \to_p S^0$, S^0 *symmetric positive definite. If* $m_n^{-1} + m_n^d n^{-1/2} \to 0$, *then*

$$\left\| \widetilde{f}^g - f \right\|_\infty \to_p 0.$$

If A.6. also holds, then

$$\left\| \widehat{f}^g - \widetilde{f}^g \right\|_\infty = O_p\left(m_n^d n^{-1}\right). \tag{5}$$

Next, I consider the asymptotic distribution of the process $\left(\widetilde{f}^g - f\right)$ at arbitrary finite sets $\lambda_1, .., \lambda_Q \in \Pi^d$. Consider $M_n(S) = m_n S$ where m_n is a deterministic sequence, and define

$$\widetilde{\nu}_n(\lambda, S) = \left(n \, m_n^{-d}\right)^{1/2} \left(\widetilde{f}^g_{M_n(S)}(\lambda) - E\left[\widetilde{f}^g_{M_n(S)}(\lambda)\right]\right),$$

where $\widetilde{f}^g_{M_n(S)}(\lambda)$ is the modified kernel estimator based on $M_n(S)$. I define $\widehat{\nu}_n(\lambda, S)$ similarly (using $\widehat{f}^g_{M_n(S)}$ instead of $\widetilde{f}^g_{M_n(S)}$). Let $\mathcal{N}$ be a compact set of symmetric positive definite matrices.

Theorem 2 *Assume A.1. to A.5. and* $\int_{\Pi^d} |K(u)|^2 \, du < \infty$, *and* $M_n(S) = m_n S$ *with* m_n *a deterministic sequence satisfying* $m_n^{-1} + m_n^d n^{-1/2} \to 0$. *Then, for any* $Q \in \mathbb{N}$ *and all finite sets* $(\lambda_1, S_1), ..., (\lambda_Q, S_Q)$ *in* $\Pi^d \times \mathcal{N}$

$$\left(\widetilde{\nu}_n(\lambda_1, S_1), .., \widetilde{\nu}_n(\lambda_Q, S_Q)\right)' \to_d \left(G(\lambda_1, S_1), , .., G(\lambda_Q, S_Q)\right)',$$

where $(G(\lambda_1, S_1),,..,G(\lambda_Q, S_Q))$ *has a Q-dimensional Gaussian with zero mean and covariance function*

$$Cov\left[G(\lambda_a, S_a), G(\lambda_b, S_b)\right] = (2\pi)^d (1+\delta(\lambda)) f(\lambda)^2 \int_{\mathbb{R}^d} k\left(S_a^{-1}u\right) k\left(S_b^{-1}u\right) du \times I(\lambda_a = \lambda_b = \lambda),$$

where $\delta(\lambda) = 1$ *when the coordinates* $\lambda_1, .., \lambda_d \in \{2\pi k : k \in \mathbb{Z}\}$ *and* $\delta(\lambda) = 0$ *otherwise. If A.7. holds, the same result is satisfied by* $\widehat{\nu}_n(\lambda, S)$.

Instead of $A.7.$, in the last statement of Theorem 2, we can use the condition $A.6$. The asymptotic distribution of $\widehat{\nu}_n(\lambda, S)$ follows from (5) and the first part of Theorem 2.

Next I ensure that the estimation of S^0 does not have an asymptotic effect on the limit distribution. Uniform weak convergence is proved applying some results from Bickel and Wichura (1971).

Theorem 3 *Under the conditions of Theorem 2, including A.7, assume that* $E\left[|\varepsilon_t|^8\right] < \infty$, *and* k *is a Lipschitz function. Then*

1. *for any* $Q \in \mathbb{N}$ *and all* $\lambda_1,,...,\lambda_Q$, *in* Π^d *there exists a Gaussian process* $G_Q(S)$ *on* $C(\mathcal{N})^Q$ *such that,*
$$(\widetilde{\nu}_n(\lambda_1, S), .., \widetilde{\nu}_n(\lambda_Q, S)) \to_d G_Q(S),$$
uniformly on $C(\mathcal{N})^Q$, *where* $G_Q(S)$ *has zero mean and covariance function as in Theorem 2. If the conditions of Theorem 1 hold, then also*
$$(\widehat{\nu}_n(\lambda_1, S), .., \widehat{\nu}_n(\lambda_Q, S)) \to_d G_Q(S),$$
uniformly on $C(\mathcal{N})^Q$.

2. *If in addition* $\int |k(u)| \|u\|_1 du < \infty$ *and* $n^{-1/2} m_n^{d+1} = O(1)$, *then for any consistent estimator* $\widehat{S}_n \to_p S^0$ *the process*
$$\widetilde{\nu}_n(\lambda) = \widetilde{\nu}_n\left(\lambda, \widehat{S}_n\right) \to_d G^0(\lambda),$$
uniformly on $C\left(\Pi^d\right)$, *where* G^0 *is Gaussian process with zero mean and covariances,*
$$Cov\left[G^0(\lambda_a), G_Q^0(\lambda_b)\right] = (2\pi)^d (1+\delta(\lambda)) f(\lambda)^2 k^2 \det\left(S^0\right) \times I(\lambda_a = \lambda_b = \lambda),$$
with $k^2 = \int_{\mathbb{R}^d} k(u)^2 du$, *and under the conditions of Theorem 1,*
$$\widehat{\nu}_n(\lambda) = \widehat{\nu}_n\left(\lambda, \widehat{S}_n\right) \to_d G^0(\lambda),$$
uniformly on $C\left(\Pi^d\right)$.

Let us consider

$$A_n(\lambda) = \frac{\left(n\, m_n^{-d}\right)^{1/2}}{\left((2\pi)^d (1+\delta(\lambda))\, k^2 \det\left(S^0\right)\right)^{1/2}} \left(\frac{\widetilde{\nu}_n(\lambda, S)}{f(\lambda)}\right).$$

By Theorem 3 and the continuous mapping theorem, for all $\phi \in C\left(\Pi^d\right)$,

$$\int_{\Pi^d} \phi(\lambda)\, A_n(\lambda)\, du \to_d \int_{\Pi^d} \phi(\lambda)\, dW(\lambda),$$

i.e., the asymptotic distribution of $\int_{\Pi^d} \phi(\lambda)\, A_n(\lambda)\, du$ is $N\left(0, \|\phi\|^2_{L_2(\Pi^d)}\right)$. An interpretation for this behavior is that the asymptotic distribution of $A_n(\lambda)$ is that of $\dot{W}$, the Gaussian white-noise generalized process on $C\left(\Pi^d\right)$.

Notice that Theorem 1 establishes uniform consistency for kernel estimators when the smoothing number has been consistently estimated from the data. Theorem 3 establishes weak convergence uniformly in $C\left(\Pi^d\right)$ when the smoothing number has been consistently estimated. Next I consider the choice of the parameter S^0, which is crucial to deal with the trade-off effects between bias and variance. Let us define the stochastic process,

$$\widetilde{\alpha}_n(\lambda, S) = \left(n\, m_n^{-d}\right)^{1/2} \left(\widetilde{f}^g_{m_n S}(\lambda) - f(\lambda)\right),$$

on $C^q\left(\Pi^d \times \mathcal{N}\right)$. Under $A.1.$ to $A.5.$, and $A.8.$, and $M_n = m_n S$, applying an argument similar to that of Hannan (1970, Th. 10, pp. 283), if $f \in C^q\left(\Pi^d\right)$

$$m_n^q \left(E\left[\widetilde{f}^g_{m_n S}(\lambda)\right] - f(\lambda)\right) \to \frac{k_q}{(2\pi)^d} S^{-q} \sum_l \|l\|^q \gamma_l e^{-il\cdot\lambda} \tag{6}$$

uniformly in $\lambda \in \Pi^d$ and $S \in \mathcal{N}$, and therefore

$$\begin{aligned} \left\| E\left[\widetilde{f}^g_{m_n S}\right] - f \right\|^2_{L_2} &= \int_{\Pi^d} \left(\frac{k_q}{m_n^q (2\pi)^d} S^{-q} \sum_l \|l\|^q \gamma_l e^{-il\cdot\lambda}\right)^2 d\lambda + o(1) \\ &= \frac{k_q^2}{m_n^{2q}} S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2 + o(1). \end{aligned}$$

Under the conditions of Theorem 3, $A.8.$ and $f \in C^q\left(\Pi^d\right)$, the continuous mapping theorem implies that

$$d\left(\int_{\Pi^d} |\widetilde{\alpha}_n(\lambda, S)|^2 d\lambda, \left(\int_{\Pi^d} G(\lambda, S)^2 d\lambda + n\, m_n^{-(2q+d)}\, k_q^2 S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2\right)\right) \to 0,$$

for any distance d that generates the weak-* topology on $C(\mathcal{N})$, where $E\left[\int_{\Pi^d} |G(\lambda, S)|^2 d\lambda\right] = \det(S)\, \kappa^2 \|f\|^2_{L_2}$. As a consequence, if we take $m_n = n^{1/(2q+d)}$, then

$$\int_{\Pi^d} |\widetilde{\alpha}_n(\lambda, S)|^2 d\lambda \to_d \int_{\Pi^d} G(\lambda, S)^2 d\lambda + k_q^2 S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2,$$

uniformly in $C(\mathcal{N})$. Therefore, I define the loss function,

$$Q(S)=\det(S)\,\kappa^2\,\|f\|_{L_2}^2+k_q^2S^{-2q}\sum_l\|l\|^{2q}\,|\gamma_l|^2$$

and define the optimal matrix S_0^* as a locally unique minimum for $Q(S)$. Similar arguments can be considered for $\widehat{f}^g_{m_nS}$. The next section considers consistent plug-in and Bootstrap estimators of the optimal smoothing number S_0^*.

Finally note that under the assumptions of Theorem 3 and $A.8.$, if m_n satisfies the condition $n\,m_n^{-(2q+d)}\to 0$, then the asymptotic bias has lrate ower than $\left(n\,m_n^{-d}\right)^{1/2}$, since

$$\left(n\,m_n^{-d}\right)^{1/2}m_n^{-q}=\left(n\,m_n^{-(2q+d)}\right)^{1/2}\to 0,$$

and therefore $\widetilde{\alpha}_n(\lambda)\to_d G^0(\lambda,)$, i.e., the asymptotic distribution of $\widetilde{f}^g$ after normalization concentrates around f without any asymptotic bias (see Hannan, 1970, pp. 288). Since the bias term tends to zero faster than the deviation term we might consider the loss function given by the integrated variance $\det(S)\,\kappa^2\,\|f\|_{L_2}^2$, and S_0^* the matrix with smallest determinant in the border of $\mathcal{N}$. Albeit for small samples, it is worthwhile to balance the bias and variance, e.g. by minimizing $Q(S)$.

4. BOOTSTRAP AND PLUG-IN ESTIMATORS

In this section I consider the Bootstrap and plug-in estimations of S_0^* for the spectral estimator $\widetilde{f}^g_{m_nS}$, but similar arguments can be considered for $\widehat{f}^g_{m_nS}$. The simplest approach is the plug-in estimation. Given a pilot estimator $\widetilde{f}^g_{m_n\widehat{S}_a}(\lambda)$, the plug-in loss function is defined by

$$Q_n^{pi}(S)=\det(S)\,\kappa^2\left\|\widetilde{f}^g_{m_n\widehat{S}_a}\right\|_{L_2}^2+k_q^2S^{-2q}\sum_l{}''\;\|l\|^{2q}\left|c_{n,l}^*\right|^2$$

where $\sum_l{}''=\sum_{i=1,..,d}\sum_{|l_i|\le g(n_i)}$. The plug-in estimator of S_0^* is given by the argument minimizing $Q_n^{pi}(S)$ on $\mathcal{N}$; i.e.

$$\widehat{S}_n^{pi}=\arg\min_{S\in\mathcal{N}}Q_n^{pi}(S).$$

Next I define a bootstrap estimator of S_0^*. I consider a Wiener random field W_u^* on $C\left(\Pi^d\right)$, which is a multiparameter analogue of a Brownian motion with covariance function $Cov\left(W_u^*,W_v^*\right)=\prod_{r=1}^d\min\{u_r,v_r\}$, and define

$$\widetilde{\alpha}_n^*(\lambda,S)=\left(n\,m_n^{-d}\right)^{1/2}\sqrt{\frac{(2\pi)^d}{n}}\int_{\Pi^d}K_{m_nS}(u-\lambda)\,I_g(u)\,dW_u^*.$$

The conditional distribution of $\alpha_n^*(\lambda, S)$ respect to the original sample is normal, with mean $E^*[\widetilde{\alpha}_n^*(\lambda, S)] = 0$ and variance is

$$Var^*[\widetilde{\alpha}_n^*(\lambda, S)] = m_n^{-d}(2\pi)^d \int_{\Pi^d} K_{m_n S}(u-\lambda)^2 I_g(u)^2 du.$$

The stochastic integral $\widetilde{\alpha}_n^*(\lambda, S)$ is determinant in the Bootstrap method.

The evaluation of $\widetilde{\alpha}_n^*(\lambda, S)$ requires the simulation of a continuous Wiener random field and the computation of a multiparameter Itô integral, which is not feasible in practice and discrete approximations are required. Thus, a discrete version can be considered, i.e.

$$\widetilde{\alpha}_n^{**}(\lambda, S) = \left(n\, m_n^{-d}\right)^{1/2} \sqrt{\frac{(2\pi)^d}{n}} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_g(\omega_{j,n}) W_{j,n}^*,$$

where $W_{j,n}^* = \prod_{r=1}^d (W^*(\omega_{j,n}) - W^*(\omega_{j-e_r,n}))$, and $I = (e_1, .., e_d)$ is the identity matrix. Note that $W_{j,n}^* = \prod_{r=1}^d \varepsilon_{n_r, j_r}$ with ε_{n_r, j_r} independently distributed $N(0, 2\pi j_r/n_r)$, for all $j \in J_n$. The expectation of $\alpha_n^{**}(\lambda, S)$ conditional to the sample is zero and the variance,

$$Var^*[\widetilde{\alpha}_n^{**}(\lambda, S)] = m_n^{-d}(2\pi)^d \left(n^{-1} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda)^2 I_g(\omega_{j,n})^2 \right).$$

The analogy with the multiparameter Itô integral is clear.

Next the Bootstrap loss function is defined, either in terms of the multiparameter Itô integral or using the discrete version, respectively given by

$$\begin{aligned} Q_n^{b*}(S) &= E^*\left[\int_{\Pi^d} |\widetilde{\alpha}_n^*(\lambda, S)|^2 d\lambda\right] + k_q^2 S^{-2q} \sum_l{}'' \|l\|^{2q} \left|c_{n,l}^*\right|^2, \\ Q_n^{b**}(S) &= E^*\left[\int_{\Pi^d} |\widetilde{\alpha}_n^{**}(\lambda, S)|^2 d\lambda\right] + k_q^2 S^{-2q} \sum_l{}'' \|l\|^{2q} \left|c_{n,l}^*\right|^2, \end{aligned}$$

where $E^*[\cdot]$ denotes the conditional expectation with respect to the data. In practice, the conditional expectation $E^*\left[|\widetilde{\alpha}_n^*(\lambda, S)|^2\right]$ can be computed by Monte Carlo methods, e.g. using the average of B realizations of $\widetilde{\alpha}_n^*(\lambda, S)$, $B^{-1}\sum_{b=1}^B \left|\widetilde{\alpha}_n^{*b}(\lambda, S)\right|^2$. Each of the values $\widetilde{\alpha}_n^{*b}(\lambda, S)$ is computed using an independent realization of the Brownian motion W_u^*. The discrete version $E^*\left[|\widetilde{\alpha}_n^{**}(\lambda, S)|^2\right]$ can be similarly computed.

The Bootstrap estimator $\widehat{S}_n^{b*}$ of S_0^*, is defined by the minimizer of $Q_n^{b*}(S)$, i.e.

$$\widehat{S}_n^{b*} = \underset{S \in \mathcal{N}}{\arg\min} Q_n^{b*}(S).$$

The Bootstrap estimator $\widehat{S}_n^{b**}$ can be defined alike, as the minimizer of $Q_n^{b**}(S)$ on $\mathcal{N}$.

The following result proves the consistency of the plug-in and the Bootstrap estimators with respect to S_0^*.

Theorem 4 *Under conditions of Theorem 3, if A.8. is satisfied,* $f \in C^q\left(\Pi^d\right)$ *and* $m_n = n^{1/(2q+d)}$, *then* $\widehat{S}_n^{pi} \to_p S_0^*$, $\widehat{S}_n^{b*} \to_p S_0^*$, *and* $\widehat{S}_n^{b**} \to_p S_0^*$.

Under the conditions of Theorem 3, if $\widehat{S}_n$ is any consistent estimator for S_0^*, then, both $\widetilde{\alpha}_n^*\left(\lambda, \widehat{S}_n\right)$ and $\widetilde{\alpha}_n^{**}\left(\lambda, \widehat{S}_n\right)$ have the same asymptotic distribution as $\left(n\, m_n^{-d}\right)^{1/2}\left(\widetilde{f}^g_{m_n S_0^*}(\lambda) - E\left[\widetilde{f}^g_{m_n S_0^*}(\lambda)\right]\right)$.

Under the conditions of Theorem 3, $A.8.$ and $f \in C^q\left(\Pi^d\right)$, we can obtain a bootstrap approximation to the distribution of $\left(n\, m_n^{-d}\right)^{1/2}\left(\widetilde{f}^g_{m_n S_0^*}(\lambda) - f(\lambda)\right)$, adding to $\widetilde{\alpha}_n^*\left(\lambda, \widehat{S}_n\right)$ a plug-in estimation of the asymptotic bias (6) scaled by $\left(n\, m_n^{-d}\right)^{1/2}$; i.e. by considering

$$\widetilde{\alpha}_n^*\left(\lambda, \widehat{S}_n\right) + \left(n\, m_n^{-d}\right)^{1/2}\left\{\frac{k_q}{m_n^q\,(2\pi)^d}\;\widehat{S}_n^{-q}\sum_l{}''\;\|l\|^q\;\; c^*_{n,l}\; e^{-il\cdot\lambda}\right\}.$$

A similar procedure can be used with $\widetilde{\alpha}_n^{**}\left(\lambda, \widehat{S}_n\right)$.

APPENDIX

Part 1. Lemmas

Lemma 1 *Assume A.1. to A.5. and* $\int_{\Pi^d}|K(u)|^2\,du < \infty$. *Then, there exists some* $C > 0$, *such that for all* $\lambda, \theta \in \Pi^d$, *the kernel estimator* $\widehat{f}^g$ *satisfies*

$$\begin{aligned} cov\left[\widehat{f}^g(\lambda), \widehat{f}^g(\theta)\right] &\leq Cn^{-2}\sum_{j\in J_n}\left|K_M(\omega_j-\lambda)\,K_M(\omega_j-\theta)\right| + \\ &\quad + Cn^{-2}\sum_{j\in J_n}\left|K_M(\omega_j-\lambda)\right|^2 + Cn^{-2}\sum_{k\in J_n}\left|K_M(\omega_k-\theta)\right|^2. \end{aligned}$$

Therefore,

$$var\left[\widehat{f}^g(\lambda)\right] \leq 3Cn^{-2}\sum_{j\in J_n}\left|K_M(\omega_j-\lambda)\right|^2 = O\left(n^{-1}\det(M)\right),$$

and the variance tends to zero for $n^{-1}\det(M_n) \to 0$.

Proof.

I follow an argument based on the proof of Theorem 1 in Robinson and Vidal-Sanz (2006). Consider the modified kernel *smoothed periodogram* estimator,

$$\widehat{f}^g(\lambda) = \frac{1}{n}\sum_{j\in J_n} K_M(\omega_{j,n}-\lambda)\, I_g(\omega_{j,n}).$$

By simplicity we will assume that $\mu = 0$, and replace X_t by $X_t - \overline{X}$, as $\overline{X}$ is $n^{1/2}$ consistent for μ under $A.3$. Clearly,

$$
\begin{aligned}
cov\left[\widehat{f}^g(\lambda), \widehat{f}^g(\theta)\right] &= n^{-2} \sum_{j \in J_n} \sum_{k \in J_n} K_M(\omega_j - \lambda) K_M(\omega_k - \theta)\, cov\left[I_g(\omega_j), I_g(\omega_k)\right] \\
&= \left\{n(2\pi)^d\right\}^{-2} \sum_{j \in J_n} \sum_{k \in J_n} K_M(\omega_j - \lambda) K_M(\omega_k - \theta) \\
&\quad \times \left\{\sum_u{}'' \sum_v{}'' cov\left[c_u^*, c_v^*\right] e^{i(v \cdot \omega_j - u \cdot \omega_k)}\right\} \qquad (7)
\end{aligned}
$$

where $\sum_u{}'' = \sum_{i=1,..,d} \sum_{|u_i| \le g(n_i)}$. The term in brackets is

$$
\begin{aligned}
& \sum_u{}'' \sum_v{}'' \frac{1}{w(u)\,w(v)} \sum_{s \in N(u)} \sum_{t \in N(v)} \left\{\gamma_{t-s-u}\gamma_{t-s+v} + \gamma_{t-s}\gamma_{t-s+v-u}\right. \\
& \left. + \, cum\left(X_s, X_{s+u}, X_t, X_{t+v}\right)\right\} e^{i(v \cdot \omega_j - u \cdot \omega_k)} \\
&= \sum_u{}'' \sum_v{}'' \frac{1}{w(u)\,w(v)} \sum_{s \in N(u)} \sum_{t \in N(v)} \left[\int_{\Pi^d} \int_{\Pi^d} f(\eta) f(\phi)\right. \\
&\quad \times \left(e^{i(t-s-u) \cdot \eta - i(t-s+v) \cdot \phi} + e^{i(t-s-u) \cdot \eta - i(t-s+v-u) \cdot \phi}\right) d\eta d\phi \\
&\quad \left. + \, k_\varepsilon \sum_{l \in \mathbb{Z}^d} \beta_{s-l}\beta_{s+u-l}\beta_{t-l}\beta_{t+v-l}\right] e^{i(v \cdot \omega_k - u \cdot \omega_j)} \qquad (8)
\end{aligned}
$$

The contribution to (7) from the first term in (8) is

$$
\begin{aligned}
& \left\{n(2\pi)^d\right\}^{-2} \sum_u{}'' \sum_v{}'' \frac{1}{w(u)\,w(v)} \int_{\Pi^d} \int_{\Pi^d} \sum_{j \in J_n} \sum_{k \in J_n} K_M(\omega_j - \lambda) K_M(\omega_k - \theta) \\
& \times e^{-iu \cdot (\eta + \omega_j) - iv \cdot (\phi - \omega_k)} \left(\sum_{s \in N(u)} \sum_{t \in N(v)} e^{i(t-s)(\eta - \phi)}\right) f(\eta) f(\phi)\, d\eta d\phi
\end{aligned}
$$

which is bounded by a constant times

$$
\left| n^{-2} \sum_u{}'' \frac{1}{w(u)} \sum_{j \in J_n} K_M(\omega_j - \lambda) K_M(\omega_j - \theta)\, e^{-iu \cdot \omega_j} \right| \qquad (9)
$$

using that $\sum_u{}''1 = w(u)$. As $w(u)^{-1} \le Kn^{-1}$ under Assumption $A.1.$, (9) is bounded by a constant times

$$
n^{-3} \sum_u{}''' \left| \sum_{j \in J_n} K_M(\omega_j - \lambda) K_M(\omega_j - \theta)\, e^{-iu \cdot \omega_j} \right| \qquad (10)
$$

where $\sum_u{}'''$ is the sum $\sum_{1-n_i\le u_i\le n_i, i=1,..,d}$, and by the triangular inequality (10) is bounded by

$$n^{-2}\sum_{j\in J_n}\left|K_M\left(\omega_j-\lambda\right)K_M\left(\omega_j-\theta\right)\right|$$

The contribution to (7) from the second term in (8) can be analogously considered, with the same order. The contribution to (7) from the third term in (8) is

$$\begin{aligned}&\left\{n\left(2\pi\right)^d\right\}^{-2}\sum_u{}''\sum_v{}''\frac{1}{w\left(u\right)w\left(v\right)}\sum_{k\in J_n}K_M\left(\omega_k-\theta\right)e^{-iu\cdot\omega_k}\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{iv\cdot\omega_j}\\&\times k_\varepsilon\sum_{s\in N(u)}\sum_{t\in N(v)}\sum_{l\in\mathbb{Z}^d}\beta_{s-l}\beta_{s+u-l}\beta_{t-l}\beta_{t+v-l}\end{aligned}$$

which is bounded by $\left(2\pi\right)^{-2d}$ times

$$\begin{aligned}&n^{-2}\sum_u{}''\sum_v{}''\frac{1}{w\left(u\right)w\left(v\right)}\left|\sum_{k\in J_n}K_M\left(\omega_k-\theta\right)e^{-iu\cdot\omega_k}\right|\left|\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{iv\cdot\omega_j}\right|\\&\times k_\varepsilon\sum_{s\in N(u)}\sum_{t\in N(v)}\sum_{l\in\mathbb{Z}^d}\left|\beta_{s-l}\beta_{s+u-l}\beta_{t-l}\beta_{t+v-l}\right|\end{aligned}$$

$$\begin{aligned}\le\;& n^{-4}\sum_v{}''\sum_u{}''\left\{\left|\sum_{k\in J_n}K_M\left(\omega_k-\theta\right)e^{iu\cdot\omega_k}\right|^2+\left|\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{iv\cdot\omega_j}\right|^2\right\}\\&\times k_\varepsilon\sum_{s\in N(u)}\sum_{t\in N(v)}\sum_{l\in\mathbb{Z}^d}\left|\beta_{s-l}\beta_{s+u-l}\beta_{t-l}\beta_{t+v-l}\right|\end{aligned}$$

that is bounded by k_ε times

$$\begin{aligned}&n^{-4}\sum_v{}''\left|\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{iv\cdot\omega_j}\right|^2\times\sum_{t\in N(v)}\sum_l\left|\beta_{t-l}\right|\sum_s\left|\beta_{s-l}\right|\sum_u\left|\beta_{t+u-l}\right|\\&+n^{-4}\sum_u{}''\left|\sum_{k\in J_n}K_M\left(\omega_k-\lambda\right)e^{iu\cdot\omega_k}\right|^2\times\sum_{s\in N(u)}\sum_l\left|\beta_{s-l}\right|\sum_t\left|\beta_{t-l}\right|\sum_v\left|\beta_{t+v-l}\right|\end{aligned}$$

and, for some constant $c>0$, it is therefore

$$\begin{aligned}\le\;& cn^{-3}\left\{\sum_u{}''\left|\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{-iu\cdot\omega_j}\right|+\sum_v{}''\left|\sum_{k\in J_n}K_M\left(\omega_k-\theta\right)e^{iv\cdot\omega_k}\right|\right\}^2\\\le\;& cn^{-2}\left|\sum_{j\in J_n}K_M\left(\omega_j-\lambda\right)e^{-iu\cdot\omega_j}\right|^2+Kn^{-2}\left|\sum_{k\in J_n}K_M\left(\omega_k-\theta\right)e^{iv\cdot\omega_k}\right|^2\end{aligned}\tag{11}$$

using that

$$\sum_{u_l=1-n_l}^{0} e^{2\pi i(k_l-j_l)/n_l} = \sum_{u_l=1}^{n_l} e^{2\pi i(k_l-j_l)n_l} = n_l \times 1\left(j_l = k_l\right), \tag{12}$$

for $1 \leq j_l, k_l \leq n_l$, the bound (11) can be expressed as

$$= 2^d c n^{-2} \sum_{j \in J_n} \left|K_M\left(\omega_j - \lambda\right)\right|^2 + 2^d c n^{-2} \sum_{k \in J_n} \left|K_M\left(\omega_k - \theta\right)\right|^2.$$

So we conclude that for some $C > 0$,

$$\begin{aligned} cov\left[\widehat{f}^g(\lambda), \widehat{f}^g(\theta)\right] &\leq Cn^{-2} \sum_{j \in J_n} \left|K_M\left(\omega_j - \lambda\right) K_M\left(\omega_j - \theta\right)\right| + \\ &\quad + Cn^{-2} \sum_{j \in J_n} \left|K_M\left(\omega_j - \lambda\right)\right|^2 + Cn^{-2} \sum_{k \in J_n} \left|K_M\left(\omega_k - \theta\right)\right|^2, \end{aligned}$$

and therefore,

$$\begin{aligned} var\left[\widehat{f}^g(\lambda)\right] &\leq 3Cn^{-2} \sum_{j \in J_n} \left|K_M\left(\omega_j - \lambda\right)\right|^2 = O\left(n^{-1} \int K_M(u-\lambda)^2 du\right) \\ &= O\left(n^{-1} \int K_M(u)^2 du\right) = O\left(n^{-1} \det(M) \int K(u)^2 du\right), \end{aligned}$$

where C does not depend on λ, n or M. ■

Part 2. Main Proofs

Proof of Theorem 1

The argument is related to that of Robinson (1991). From Assumption $A.5.$, there exists a continuous spectral density $f \in C\left(\Pi^d\right)$, and it obeys the Lipschitz condition $f \in Lip(\alpha)$ for any $\alpha > 1/2$. Let $\mathcal{N}$ be the closure of a neighborhood of matrix S^0. Then,

$$\Pr\left\{\left\|\widetilde{f}^g - f\right\|_\infty > \eta\right\} \leq \Pr\left\{\max_{S \in \mathcal{N}} \max_{\lambda \in \Pi^d} \left|\widetilde{f}^g_{M_n(S)}(\lambda) - f(\lambda)\right| > \eta\right\} + \Pr\left(\widehat{S}_n \notin \mathcal{N}\right),$$

where the last term in the right hand side tends to zero.

We can express $\widetilde{f}^g(\lambda) - f(\lambda) = \sum_{j=1}^2 \widetilde{b}^j_M(\lambda)$ where

$$\begin{aligned} \widetilde{b}^1_M(\lambda) &= E\left[\widetilde{f}^g_M(\lambda)\right] - f(\lambda) = \int_{\Pi^d} K_M(u-\lambda)\left(E\left[I_g(u)\right] - f(\lambda)\right) du, \\ \widetilde{b}^2_M(\lambda) &= \widetilde{f}^g_M(\lambda) - E\left[\widetilde{f}^g_M(\lambda)\right] = \int_{\Pi^d} K_M(u-\lambda)\left(I_g(u) - E\left[I_g(u)\right]\right) du \\ &= (2\pi)^{-d} \sum_u{}'' \, k\left(M^{-1}u\right)\left(c^*_u - \gamma_u\right) e^{-iu'\lambda}. \end{aligned}$$

For the bias term, I use that

$$\begin{aligned}
\max_{M,\lambda}\left|\widetilde{b}_M^1(\lambda)\right| &\leq \max_{M,\lambda}\left|\int_{\Pi^d} K_M(u-\lambda)\left(E\left[I_g(u)\right]-f(\lambda)\right)du\right| \\
&\quad +\max_{M,\lambda}\left|\int_{\Pi^d} K_M(u-\lambda)\left(f(u)-f(\lambda)\right)du\right| \\
&\leq \max_{M,\lambda}\left|\int_{\Pi^d} K_M(u-\lambda)\left(f(u)-f(\lambda)\right)du\right| \\
&\quad +\left\|E\left[I_g\right]-f\right\|_\infty \max_{M,\lambda}\int_{\Pi^d}\left|K_M(u-\lambda)\right|du. \qquad (13)
\end{aligned}$$

Since f is periodic, using an argument similar to the Korovkin Theorem, the first term in (13) is

$$\max_M\left\|\int K_M(u-\lambda)\left(f(u)-f(\lambda)\right)du\right\|_\infty = O\left(\max_{S\in\mathcal{N}}\omega\left(f,\delta_M\right)\right)=O\left(\max_{S\in\mathcal{N}}\delta_M^{-1/2}\right),$$

where $\delta_M = \int\|u\|\left|K_M(u)\right|du = O\left(\|M\|^{-1}\right) = O\left(m_n^{-1}\|S\|^{-1}\right) \rightarrow_p 0$ if $m_n \rightarrow_p \infty$. Higher convergence rates can be established using higher order kernels as previously explained.

The second term in (13) is

$$O\left(\left\|E\left[I_g\right]-f\right\|_\infty\int_{\Pi^d}\left|K(u)\right|du\right)=o\left(n^{-1/2}\right),$$

under $A.4.$, since

$$\left\|E\left[I_g\right]-f\right\|_\infty \leq (2\pi)^d \sum_{\substack{l\in\mathbb{Z}:\exists r\in\{1,\ldots,d\} \\ with\ \left|l_r\right|>g(n_r)}}\left|\gamma_l\right| = o\left(n^{-1/2}\right),$$

see the proof of Theorem 1 of Robinson and Vidal-Sanz (2006).

Recall that $n(l)\left(c_{n,l}^*-\gamma_l\right)$ is a sum of $n(l)$ terms $\left(X_tX_{n-l+t}-E\left[X_tX_{n-l+t}\right]\right)$, and consider also the classical result that for all l_1, l_2

$$\begin{aligned}
\lim_{n\rightarrow\infty} n(l)\,E\left[\left(c_{n,l_1}^*-\gamma_{l_1}\right)\left(c_{n,l_2}^*-\gamma_{l_2}\right)\right] &= 2(2\pi)^d\int_{\Pi^d} e^{il_1\cdot\lambda}e^{il_2\cdot\lambda}f(\lambda)^2\,d\lambda \\
&\quad +\kappa_\varepsilon(2\pi)^{2d}\int_{\Pi^d} e^{il_1\cdot\lambda}f(\lambda)\,d\lambda\int_{\Pi^d} e^{il_2\cdot\lambda}f(\lambda)\,d\lambda
\end{aligned}$$

uniformly in l. Since $\sup_l\left|e^{il\cdot\lambda}\right|\leq 1$ for all $\lambda\in\Pi^d$, then $n(l)^{1/2}\left(c_{n,l}^*-\gamma_l\right)=O_p(1)$ uniformly in l when $f(\lambda)^2$ is integrable. Next, under $A.2.$,

$$\sup_{\{l:|l_r|\leq g(n_r),r=1,..,d\}}\frac{1}{\prod_{r=1}^d\left(n_r-\left|l_r\right|\right)}\leq\frac{1}{\prod_{r=1}^d\left(n_r-g(n_r)\right)}\leq\frac{1}{(1-c_2)}\frac{1}{\prod_{r=1}^d n_r}=\frac{1}{(1-c_2)}\frac{1}{n},$$

as $\left(n_r-g(n_r)\right)\geq(1-c_2)\,n_r$, with $c_2\in(0,1)$. Then

$$\sup_{\{l:|l_r|\leq g(n_r),r=1,..,d\}}\left|c_{n,l}^*-\gamma_l\right|=O_p\left(n^{-1/2}\right).$$

Furthermore, by dominated convergence arguments

$$E\left[n^{1/2}\sup_{\{l:|l_r|\leq g(n_r),r=1,..,d\}}\left|c^*_{n,l}-\gamma_l\right|\right]=O\left(1\right),$$

as $\left|c^*_{n,l}-\gamma_l\right|\leq\left|c^*_{n,0}\right|+|\gamma_0|$ with $E\left[\left|c^*_{n,0}\right|\right]<\infty$.

Regarding the variance term,

$$\begin{aligned}
\max_{S\in\mathcal{N}}\max_{\lambda\in\Pi^d}\left|\widetilde{b}^2_M\left(\lambda\right)\right| &\leq (2\pi)^{-d}\max_{S\in\mathcal{N}}\sum_u{}'' \left|k\left(M^{-1}u\right)\right|\left|c^*_u-\gamma_u\right| \\
&= O_p\left(n^{-1/2}\max_{S\in\mathcal{N}}\sum_u{}'' \left|k\left(\frac{S^{-1}u}{m_n}\right)\right|\right) \\
&= O_p\left(n^{-1/2}m_n^d\max_{S\in\mathcal{N}}\int_{\mathbb{R}^d}\left|k\left(S^{-1}z\right)\right|dz\right) \\
&= O_p\left(n^{-1/2}m_n^d\int_{\mathbb{R}^d}\left|k\left(z\right)\right|dz\max_{S\in\mathcal{N}}\det\left(S\right)\right)
\end{aligned}$$

using that $m_n^{-1}\to 0$. The last expression is negligible using that $\det\left(S\right)<\epsilon$ on $\mathcal{N}$ for some $\epsilon>0$, as $\mathcal{N}$ is a compact set of positive definite matrices, and using the condition $m_n^d n^{-1/2}\to 0$. I have used that k is continuous and integrable under $A.5$.

Next, I consider $\left|\widehat{f}^g_M-\widetilde{f}^g_M\right|$, defining $o_n=(n_1,...,n_r)$ and applying an argument analogous to that of Hannan (1973),

$$\begin{aligned}
\widehat{f}^g_M\left(\lambda\right) &= \frac{1}{n}\sum_{j\in J_n}K_M\left(\omega_{j,n}-\lambda\right)I^g_*\left(\omega_{j,n}\right) \\
&= \widetilde{f}^g_M\left(\lambda\right)+(2\pi)^{-d}\sum_l{}''k\left(S^{-1}\frac{l}{m_n}\right)c^*_{l\pm o_n}
\end{aligned}$$

where $c^*_{l\pm o_n}=0$ for $l=0$ and $c^*_{l\pm o_n}=c^*_{o_n-l}$ for $l\neq 0$. Then, under $A.6.$,

$$\max_{S\in\mathcal{N}}\max_{\lambda\in\Pi^d}\left|\widehat{f}^g_M\left(\lambda\right)-\widetilde{f}^g_M\left(\lambda\right)\right|\leq(2\pi)^{-d}\max_{S\in\mathcal{N}}\sum_l{}''\left|k\left(S^{-1}\frac{l}{m_n}\right)\right|\left|c^*_{l\pm o_n}\right|,$$

using that for $|l_r|\leq g\left(n_r\right), r=1,..,d$ the covariance $c^*_{(o_n-l)}$ is a sum of $\prod_{r=1}^d|l_r|$ terms of the form $X_tX_{t+o_n-l}$, all with with finitely bounded mean, divided by $n\left(o_n-l\right)=\prod_{r=1}^d|l_r|$. Under $A.2.$, for $|l_r|\leq g\left(n_r\right), r=1,..,d$

$$n\left(o_n-l\right)\leq\prod_{r=1}^d g\left(n_r\right)\leq c_2\prod_{r=1}^d n_r=c_2n,$$

so that the aliasing effect is avoided. Therefore,

$$\begin{aligned}
\max_{S\in\mathcal{N}}\left\|\widehat{f}^g_M(\lambda)-\widetilde{f}^g_M(\lambda)\right\|_\infty &= O_p\left(n^{-1}\max_{S\in\mathcal{N}}\sum_l{}''\prod_{r=1}^d|l_r|\left|k\left(S^{-1}\frac{l}{m_n}\right)\right|\right)\\
&= O_p\left(n^{-1}m_n^d\max_{S\in\mathcal{N}}\int_{\mathbb{R}^d}\prod_{r=1}^d|u_r|\left|k\left(S^{-1}u\right)\right|du\right),
\end{aligned}$$

that tends to zero if $n^{-1/2}m_n^d\to 0$.

Proof of Theorem 2

First, I prove the result for $\widetilde{f}^g$. Define $A_{S,m}=\left\{u:\left|\left(S^{-1}u\right)_r\right|<m, r=1,...,d\right\}$. Under $A.7.$ we can express

$$\widetilde{f}^g_{M_n(S)}(\lambda)-E\left[\widetilde{f}^g_{M_n(S)}(\lambda)\right]=(2\pi)^{-d}\sum_{l\in A_{S,m_n}}k\left(\frac{S^{-1}l}{m_n}\right)(c_l^*-\gamma_l)\,e^{-il\cdot\lambda}.$$

For any $Q\in\mathbb{N}$ and any $(\lambda_1,S_1),...,(\lambda_Q,S_Q)$ in $\Pi^d\times\mathcal{N}$, I apply the Cramer-Wold device, consider $(\delta_1,...,\delta_Q)'\in\mathbb{R}^d/\{0\}$,

$$\begin{aligned}
V_n(m_n)&=m_n^{-d/2}\sum_{q=1}^Q\delta_q\left(\widetilde{f}^g_{M_n(S_q)}(\lambda_q)-E\left[\widetilde{f}^g_{M_n(S_q)}(\lambda_q)\right]\right)\\
&=(2\pi)^{-d}\sum_{q=1}^Q\delta_q m_n^{-d/2}\sum_{l\in A_{S_q,m_n}}k\left(S_q^{-1}\frac{l}{m_n}\right)e^{-il\cdot\lambda_q}(c_l^*-\gamma_l).
\end{aligned}$$

Next I use the Bernstein Lemma. For each fixed m, applying a Central Limit Theorem argument in Theorem 1 of Robinson and Vidal-Sanz (2006), and the delta method, we have

$$n^{1/2}V_n(m)\to_d N(0,\Omega_m),$$

where

$$\begin{aligned}
\Omega_m &= (2\pi)^{-d}\sum_{q=1}^Q\sum_{q'=1}^Q\delta_q\delta_{q'}\sum_{l\in A_{S_q,m}}\sum_{l'\in A_{S_{q'},m}}\left(m^{-d}k\left(S_q^{-1}\frac{l}{m}\right)k\left(S_{q'}^{-1}\frac{l'}{m}\right)e^{-i(l-l')\cdot\lambda_q}\right)\\
&\quad\times(2\pi)^{2d}(1+\delta(\lambda_q))\,f(\lambda_q)^2\times I\left(\lambda^q=\lambda^{q'}\right).
\end{aligned}$$

The key argument in Robinson and Vidal-Sanz (2006) is that

$$c_l^*-\gamma_l=\frac{1}{\prod_{r=1}^d(n_r-|l_r|)}\sum_j\sum_k\beta_j\beta_k\left\{\sum_{t\in N(l)}\varepsilon_{t-j}\varepsilon_{t+l-k}-\sigma^2I(j=k-l)\right\}, \tag{14}$$

and then, after reordering the expression as a triangular array, a martingale differences central limit theorem is applied.

Since

$$\begin{aligned}\Omega_m \quad \to \quad \Omega &= \sum_{q=1}^{Q}\sum_{q'=1}^{Q}\delta_q\delta_{q'}\int_{\mathbb{R}^d} k\left(S_q^{-1}u\right)k\left(S_{q'}^{-1}u\right)du \\ &\quad \times (2\pi)^d\left(1+\delta\left(\lambda_q\right)\right)f\left(\lambda_q\right)^2 \times I\left(\lambda^q=\lambda^{q'}\right)\end{aligned}$$

as $m\to\infty$, the result follows by Bernstein's Lemma, after proving that for all $\varepsilon>0$,

$$\lim_{n\to\infty} E\left[\left|V_n\left(m\right)-V_n\left(m_n\right)\right|^2\right]=O\left(\varepsilon\right),$$

for any large enough fixed m. This is immediate, since $V_n\left(m\right)-V_n\left(m_n\right)$ has zero mean and variance given by,

$$\begin{aligned}&(2\pi)^{-2d}\sum_{q=1}^{Q}\sum_{q'=1}^{Q}\delta_q\delta_{q'}\left(m^{-d}\sum_{l\in A_{S_q,m}}k\left(S_q^{-1}\frac{l}{m}\right)-m_n^{-d}\sum_{l\in A_{S_q,m_n}}k\left(S_q^{-1}\frac{l}{m_n}\right)\right)\\ &\times\left(m^{-d}\sum_{l'\in A_{S_{q'},m}}k\left(S_{q'}^{-1}\frac{l'}{m}\right)-m_n^{-d}\sum_{l'\in A_{S_{q'},m_n}}k\left(S_{q'}^{-1}\frac{l'}{m_n}\right)\right)\\ &\times e^{-i\left(l\cdot\lambda_q-l'\cdot\lambda_{q'}\right)}Cov\left[c_l^*,c_{l'}^*\right].\end{aligned}$$

Using an argument analogous to Lemma 1, the variance $Var\left[\left|V_n\left(m\right)-V_n\left(m_n\right)\right|\right]$ is bounded by a finite sum of terms

$$\begin{aligned}&O\left(n^{-1}\left|m^{-d}\sum_{l\in A_{S,m}}k\left(S^{-1}\frac{l}{m}\right)-m_n^{-d}\sum_{l'\in A_{S,m}}k\left(S^{-1}\frac{l'}{m_n}\right)\right|^2\right)\\ =\;&O\left(n^{-1}\left|m^{-d}\sum_{l}k\left(S^{-1}\frac{l}{m}\right)-\int k\left(S^{-1}u\right)du\right|^2\right)\\ &+O\left(n^{-1}\left|\int k\left(S^{-1}u\right)du-m_n^{-d}\sum_{l'}k\left(S^{-1}\frac{l'}{m_n}\right)\right|^2\right)\\ =\;&O\left(n^{-1}\varepsilon^2\right)+O\left(n^{-1}m_n^{-2d}\right)\to 0,\end{aligned}$$

where ε is arbitrarily small when m is large, and I have used that if $k\left(u\right)$ has an integrable derivative then $\left|\int k\left(u\right)du-m_n^{-d}\sum_l k\left(l/m_n\right)\right|=O\left(m_n^{-d}\right)$.

Next I consider the smoothed periodogram. I start considering a fixed $M=mS$, and obtaining the limit distribution of $n^{1/2}\left(\widehat{f}_M^g-E\left[\widehat{f}_M^g\right]\right)$. Let $K_{ML}\left(u\right)$ be the the L-order Fejer approximation to $K_M(u)$, i.e.

$$K_{ML}(u) = \sum_{\ell \in A_L} \left(1 - \frac{|\ell|}{L}\right) k_{M,\ell} e^{-i\ell \cdot u},$$

where $k_{M\ell} = \int_{\Pi^d} K_M(u) e^{i\ell \cdot u} du = k\left(M^{-1}\ell\right)$. For a large enough L, we can ensure that $\|K_M - K_{ML}(u)\|_\infty < \varepsilon$. If we define

$$\begin{aligned}
r_M(\lambda) &= n^{-1} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda)(I_g(\omega_{j,n}) - E[I_g(\omega_{j,n})]) \\
r_{ML}(\lambda) &= n^{-1} \sum_{j \in J_n} K_{ML}(\lambda)(I_g(\omega_{j,n}) - E[I_g(\omega_{j,n})])
\end{aligned}$$

then $r(\lambda) - r_{ML}(\lambda)$ has zero mean and variance

$$n^{-1} \sum_{j \in J_n} \sum_{l \in J_n} \overline{K}_L(\omega_{j,n} - \lambda) \overline{K}_L(\omega_{l,n} - \lambda) \left\{ \sum_u {}'' \sum_v {}'' cov\left[c_\ell^*, c_\ell^*\right] e^{-i(u-v)\cdot \omega_{j,n}} \right\}$$

where $\overline{K}_{ML}(\lambda) = K_M(\lambda) - K_{LM}(\lambda)$. This variance is arbitrarily small by an argument analogous to that of Lemma 1. By the Bernstein Lemma it suffices to obtain the asymptotic distribution of $n^{1/2} r_{ML}(\lambda)$ in the sense of finite dimensional projections. Using (12),

$$\begin{aligned}
r_{ML}(\lambda) = n^{-1} & \sum_{j \in J_n} K_{ML}(\omega_{j,n} - \lambda)(I_g(\omega_{j,n}) - E[I_g(\omega_{j,n})]) \\
= \; & n^{-1} \sum_{j \in J_n} \left(\sum_{\ell \in A_L} \left(1 - \frac{|\ell|}{L}\right) k\left(M^{-1}\ell\right) e^{-i\ell \cdot (\omega_{j,n} - \lambda)} \right) \\
& \times \left\{ (2\pi)^{-d} \sum_u {}'' (c_u^* - \gamma_u) e^{-iu \cdot \omega_{j,n}} \right\} \\
= \; & (2\pi)^{-d} \sum_{\ell \in A_L} \left(1 - \frac{|\ell|}{L}\right) k\left(M^{-1}\ell\right)(c_\ell^* - \gamma_\ell) e^{-i\ell \cdot (\omega_{j,n} - \lambda)}
\end{aligned}$$

for n large enough, because then $L + g(n_r) < n_r$ for all $r = 1, ..., d$, and aliased terms do not contribute.

Expressing $M = mS$, then for any $Q \in \mathbb{N}$ and any $(\lambda_1, S_1), ..., (\lambda_Q, S_Q)$ in $\Pi^d \times \mathcal{N}$, I apply the Cramer-Wold device, consider $(\delta_1, ..., \delta_Q)' \in \mathbb{R}^d / \{0\}$,

$$\begin{aligned}
n^{1/2} V_L(m) = \; & n^{1/2} m^{-d/2} \sum_{q=1}^{Q} \delta_q \times r_{(mM_q)L}(\lambda_q) \\
= \; & m^{-d/2} (2\pi)^{-d} \sum_{q=1}^{Q} \delta_q \sum_{\ell \in A_L} \left(1 - \frac{|\ell|}{L}\right) k\left((mS_q)^{-1} \ell\right) e^{-i\ell \cdot (\omega_{j,n} - \lambda)} n^{1/2} (c_\ell^* - \gamma_\ell).
\end{aligned}$$

Since $n^{1/2} V_L(m)$ converges weakly to a normal $N(0, \Omega_{Lm})$, where $\Omega_{Lm} \to \Omega$ when $L, m \to \infty$, the result follows by an analogous argument to the previous case.

Proof of Theorem 3

Theorem 2 has established the weak convergence of finite-dimensional projections. Therefore, we only need to prove the Tightness in $C(\mathcal{N})$, and the theorem will follows from Prohorov´s Theorem. To show tightness, applying a Bickel and Wichura (1971) criterion and Cauchy-Schwartz inequality, we require bounds on fourth moments of differences,

$$E\left[\left|\widetilde{\nu}_n(\lambda, S) - \widetilde{\nu}_n(\lambda, S')\right|^4\right] \leq c\left\|S_2 - S_1\right\|^{4\alpha},$$

for some constant c, and $\alpha > 0$, where $\|S\| = \|vec(S)\|_\infty$, with $vec(S) = (S_1', ..., S_d')'$ for all square matrix $S = (S_1, .., S_d)$. Therefore $\|S_2 - S_1\|$ is bounded by the Lebesgue measure of the interval defined by $\{vec(S_1), vec(S_2)\}$.

Since $n^{-1/2}\sup_{l:|l_r|\leq g(n_r)}\left|c_{n,l}^* - \gamma_l\right| = O_p(1)$ and $\left|c_{n,l}^* - \gamma_l\right| \leq \left|c_{n,0}^*\right| + |\gamma_0|$ with $E\left[\left|c_{n,0}^*\right|^4\right] < \infty$, by dominated convergence arguments it is satisfied that $n^{-2}E\left[\sup_{l:|l_r|\leq g(n_r)}\left|c_{n,l}^* - \gamma_l\right|^4\right] = O(1)$. Therefore,

$$E\left[\left|\widetilde{\nu}_n(\lambda, S_2) - \widetilde{\nu}_n(\lambda, S_1)\right|^4\right] \leq cE\left[\left\{\sum_{r=1}^{d}\sum_{|l_r|\leq g(n_r)}\left|k\left(\frac{S_2^{-1}l}{m_n}\right) - k\left(\frac{S_1^{-1}l}{m_n}\right)\right|\left|c_{n,l}^* - \gamma_l\right|\right\}^4\right]$$

$$= O\left(n^{-2}\left\{\sum_{r=1}^{d}\sum_{|l_r|\leq g(n_r)}\left|k\left(\frac{S_2^{-1}l}{m_n}\right) - k\left(\frac{S_1^{-1}l}{m_n}\right)\right|\right\}^4\right),$$

$$= O\left(\left\{n^{-1/2}m_n^d\int\left|k\left(S_2^{-1}u\right) - k\left(S_1^{-1}u\right)\right|du\right\}^4\right)$$

and since k is Lipschitz with compact support A,

$$\begin{aligned}\int\left|k\left(S_2^{-1}u\right) - k\left(S_1^{-1}u\right)\right|du &\leq c'\int_A\left|\left(S_2^{-1} - S_1^{-1}\right)u\right|^\alpha du \\ &\leq c'\left\|S_1 - S_2\right\|^\alpha\int_A\left|\left(S_1S_2\right)^{-1}u\right|^\alpha du \\ &\leq c''\left\|S_1 - S_2\right\|^\alpha,\end{aligned}$$

using that $\left(S_2^{-1} - S_1^{-1}\right) = (S_1 - S_2)(S_1S_2)^{-1}$ for S_2, S_1 symmetric, and $\det(S_1S_2) \geq \epsilon$ on $\mathcal{N}$, for some $\epsilon > 0$. The result follows using that $n^{-(1+\varepsilon)/2}m_n^d \to 0$. The uniform weak convergence for $\widehat{\nu}_n(\lambda, S)$ follows from the uniform weak convergence of $\widetilde{\nu}_n(\lambda, S)$, and (5).

Finally, for $\widetilde{\nu}_n(\lambda_1) = \widetilde{\nu}_n\left(\lambda, \widehat{S}_n\right)$ I consider the uniform weak convergence in Π^d. To establish tightness, it suffices that

$$E\left[\left|\widetilde{\nu}_n(\lambda_1) - \widetilde{\nu}_n(\lambda_2)\right|^4\right] \leq c\left\|\lambda_1 - \lambda_2\right\|^4,$$

Since

$$E\left[\left|\widetilde{\nu}_n(\lambda_1)-\widetilde{\nu}_n(\lambda_2)\right|^4\right]\le cE\left[\left\{\sum_{r=1}^{d}\sum_{|l_r|\le g(n_r)}\left|k\left(\frac{\widehat{S}_n^{-1}l}{m_n}\right)\right|\left|c_{n,l}^*-\gamma_l\right|\left|\cos(\lambda_1\cdot l)-\cos(\lambda_2\cdot l)\right|\right\}^4\right],$$

$$= O\left(n^{-2}\left|\sum_{r=1}^{d}\sum_{|l_r|\le g(n_r)}\left|k\left(\frac{\widehat{S}_n^{-1}l}{m_n}\right)\right|\left|\cos(\lambda_1\cdot l)-\cos(\lambda_2\cdot l)\right|\right|^4\right)$$

$$= O\left(\left|n^{-1/2}\sum_{r=1}^{d}\sum_{|l_r|\le g(n_r)}\left|k\left(\frac{\widehat{S}_n^{-1}l}{m_n}\right)\right|\left|(\lambda_1-\lambda_2)\cdot l\right|\right|^4\right)$$

$$= O\left(\left|n^{-1/2}m_n\sum_{r=1}^{d}\sum_{|l_r|\le g(n_r)}\left|k\left(\frac{\widehat{S}_n^{-1}l}{m_n}\right)\right|\left\|\frac{l}{m_n}\right\|_1\left\|\lambda_1-\lambda_2\right\|_\infty\right|^4\right),$$

where I have used that cos is Lipschitz (as its derivative is bounded). Using the fact if g has integrable first order derivatives then (using the multivariate Euler-Maclaurin formulae)

$$\sup_{S\in\mathcal{N}}\left|m_n^{-d}\sum_l g\left(S^{-1}l/m_n\right)-\int g\left(S^{-1}u\right)du\right| = O\left(m_n^{-d}\sup_{S\in\mathcal{N}}\left\{\max_{j=1,,.d}\int\left|\frac{\partial}{\partial u_j}g\left(S^{-1}u\right)\right|du\right\}\right),$$

and $\widehat{S}_n\to_p S^0$, then

$$\begin{aligned}
E\left[\left|\widetilde{\nu}_n(\lambda_1)-\widetilde{\nu}_n(\lambda_2)\right|^4\right] &= O\left(\left\{n^{-1/2}m_n^{d+1}\int\left|k\left(S_0^{-1}u\right)\right|\|u\|_1\,du\right\}^4\|\lambda_1-\lambda_2\|_\infty^4\right)\\
&= O\left(\left\{n^{-1/2}m_n^{d+1}\right\}^4\|\lambda_1-\lambda_2\|_\infty^4\right)\\
&= O\left(\|\lambda_1-\lambda_2\|_\infty^4\right),
\end{aligned}$$

as $\sup_{S\in\mathcal{N}}\int\left|k\left(S^{-1}u\right)\right|\|u\|_1\,du<\infty$, and $n^{-1/2}m_n^{d+1}=O(1)$.

Proof of Theorem 4

Let $\mathcal{N}$ be the closure of a neighborhood of S_0^*. It is straightforward that

$$\sup_{S\in\mathcal{N}}\left|Q_n^{pi}(S)-Q(S)\right|\to_p 0$$

and the consistency of the plug-in estimator follows by an standard application of consistency theorem for extremum estimators. Regarding the Bootstrap estimator,

$$Var^*\left[\widetilde{\alpha}_n^*(\lambda,S)\right]=m_n^{-d}(2\pi)^d\int_{\Pi^d}K_{m_nS}(u-\lambda)^2 I_g(u)^2\,du$$

Applying Fubini's Theorem,

$$\begin{aligned}
&\int_{\Pi^d} E\left[\left|\widetilde{\alpha}_n^*(\lambda,S)\right|^2\right] d\lambda = m_n^{-d}(2\pi)^d \int_{\Pi^d}\int_{\Pi^d} K_{m_nS}(u-\lambda)^2\, d\lambda \\
&\times \left\{\sum_l{}'' \sum_k{}'' c_l^* c_k^* e^{i(l-k)\cdot u}\right\} du \\
= \quad & (2\pi)^d m_n^{-d} \int_{\Pi^d} K_{m_nS}(\lambda)^2\, d\lambda \int_{\Pi^d}\left(\sum_l{}'' \sum_k{}'' c_l^* c_k^* e^{i(l-k)\cdot u}\right) du \\
= \quad & m_n^{-d}\int_{\Pi^d} K_{m_nS}(\lambda)^2\, d\lambda \sum_l{}'' \left|c_l^*\right|^2 = \det(S)\,\kappa^2 \sum_l{}'' \left|c_l^*\right|^2 \\
\to \quad & {}_p \det(S)\,\kappa^2\left(\sum_l \left|\gamma_l\right|^2\right) = \det(S)\,\kappa^2 \int_{\Pi^d} f(u)^2\, du,
\end{aligned}$$

uniformly in $S \in \mathcal{N}$. The convergence in probability can be proved analogously to the proof of second part of Theorem 1. Since the only stochastic term $\sum_l{}'' \left|c_l^*\right|^2$ is independent of S, this convergence is uniform on in $S \in \mathcal{N}$. Therefore

$$\sup_{S\in\mathcal{N}} \left|Q_n^{b*}(S) - Q(S)\right| \to_p 0$$

and the consistency $\widehat{S}_n^{b*} \to_p S_0^*$ follows. The consistency of $\widehat{S}_n^{b**}$ can be analogously established.

REFERENCES

[1] Abramenko, V; V. Yurchyshyn; H. Wang and P. R. Goode (2001): Magnetic power spectra derived from ground and space measurements of the solar magnetic fields. *Solar Physics*, 201, 225-240.

[2] Agterberg, F. P. (1967): Computer techniques in Geology. *Earth Science Reviews*, 3, 47-77.

[3] Barry, R. G. and A. Perry (1973): *Synoptic climatology.* Methuen, London.

[4] Beltrao, K. I. and P. Bloomfield (1987): Determining the bandwith of a kernel spectrum estimate. *Journal of Time Series Analysis*, **15**, 383-400.

[5] Bickel, P. J. and M. J. Wichura (1971): Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics*, **42**, 1656-1670.

[6] Bolthausen, E. (1982): On the CLT for stationary mixing random fields. *Annals of Probability*, **10**, 1047-1050.

[7] Brillinger, D. R. (1970): The frequency analysis of relations between stationary spatial series. In: *Proceedings of the 12th biennial seminar on Time series, Stochastic Processes, Convexity and Combinatorics*, R. Pyke ed. *Canadian Math. Congress*, Montreal, pp. 39-81.

[8] Brillinger, D. R. (1981): *Time Series, Data Analisis and Theory*, Holden Day Inc., San Francisco.

[9] Dalhaus, R. and H. Künsch (1987): Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, **74**, 877-882.

[10] Delgado, M. and J. M. Vidal-Sanz (2001): Averaged Singular Integral Estimation as Bias Reduction Technique. *Journal of Multivariate Analysis*, **80**, 127-137.

[11] Doukhan, P. (1994): *Mixing. Properties and Examples.* Springer Verlag, New York.

[12] Frank, J. and W. Härdle (1992): On bootstrapping kernel spectral estimates. *Annals of Statistics*, **20**, 121-145.

[13] Guyon, X. (1982): Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, **69**, 95-105.

[14] Hannan, E. J. (1970): *Multiple Time Series.* John Wiley & Sons, New York.

[15] Hannan, E. J. (1973): The asymptotic theory of linear time-series models. *Journal of Applied Probability*, **10**, 130-145.

[16] Heyde, C. C. and R. Gay (1993): Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and Their Applications*, 45, 169-182.

[17] Ivanov, A. V. and N. N. Leonenko (1986): *Statistical Analysis of Random Fields.* Mathematics and its applications, Soviet subseries. Kluwer Academic Pubishers, Dordrecht, The Netherlands.

[18] P. P. Korovkin (1960): *Linear operators and approximation theory*, Hindustan Publ. Corp., Delhi, India.

[19] Leonenko, N. (1999): *Limit theorems for random fields with singular spectrum.* Series: Mathematics and its applications, vol. **465**. Kluwer Academic Publishers, Dordrecht, Neetherlands.

[20] Politis, D.N. and Romano, J. P. (1996): On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*, 51, 41-53.

[21] Priestley, M. B. (1964): The analysis of a two-dimensional stationary process with discontinuous spectra. *Biometrika*, 51, 195-217.

[22] Reshaw, E. (1984): Competition experiments in a plant monoculture: an analysis based on two dimensional spectra. *Biometrics*, 40, 717-728.

[23] Reshaw, E. and E. D. Ford (1983): The interpretation of process from pattern using two dimensional spectral analysis: methods and problems of interpretation. *Applied Statistics*, 32, 51-63.

[24] Ripley, B. D. (1988): *Statistical Inference for Spatial Processes.* An essay awarded the Adams Prize of the University of Cambridge. Cambridge University Press, Cambridge.

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

[26] Robinson, P. M. (2007): Nonparametric Spectrum Estimation for Spatial Data. Special Issue in honour of Madan L. Puri, *Journal of Statistical Planning and Inference*, 137, 1024-1034.

[27] Robinson, P. M. and J. M. Vidal-Sanz (2006): Modified Whittle Estimation of Multilateral Models on a Lattice. *Journal of Multivariate Analysis*, 97, 1090-1120.

[28] Rosenblatt (1985): *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

[29] Rozanov, Y. A. (1967): *Stationary Random Process.* Holden day, San Francisco.

[30] Rudin, W. (1974): *Real and Complex Analysis.* (2nd ed.). New York, Mc Graw Hill.

[31] Vidal-Sanz, J. M. (2005): Pointwise Universal Consistency of Nonparametric Linear Estimators. *Bernoulli*, 11 (4), 1-15.

[32] Whittle, P. (1954): On stationary processes in the plane. *Biometrika*, **41**, 434-449.

[33] Zygmund, A. (1959): *Trigonometric series*, second edition, vols. I and II Combined, Cambridge Mathematical Library, Cambridge University Press, Cambridge.

[34] Žurbenko, I. G. (1986): *The Spectral Analysis of Time Series.* Amsterdam, North Holland.