Biometrika (2004), **91**, 1, *pp.* 211–218

Printed in Great Britain

Contiguity of the Whittle measure for a Gaussian time series

By NIDHAN CHOUDHURI

Department of Statistics, Case Western Reserve University, Cleveland, Ohio 44106, U.S.A.

nidhan@nidhan.cwru.edu

SUBHASHIS GHOSAL

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

ghosal@stat.ncsu.edu

and ANINDYA ROY

Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland 21250, U.S.A.

anindya@math.umbc.edu

Summary

For a stationary time series, Whittle constructed a likelihood for the spectral density based on the approximate independence of the discrete Fourier transforms of the data at certain frequencies. Whittle's likelihood has been widely used in the literature for constructing estimators. In this paper, we show that, for a Gaussian time series, the Whittle measure is mutually contiguous with the actual distribution of the data. As a consequence, most asymptotic properties of estimators and test statistics derived under the Whittle measure can be carried over to the actual distribution.

Some key words: Consistency; Contiguity; Discrete Fourier transform; Periodogram; Spectral density; Whittle likelihood.

1. Introduction

Let $X = (X_1, \ldots, X_n)'$ be the vector of n observations from a zero-mean stationary time series with autocovariance function $\gamma(.)$ and the spectral density f. Whittle (1957, 1962) observed that the discrete Fourier transform of the data at certain selected frequencies can be approximated by uncorrelated zero-mean Gaussian random variables whose variances are easily expressible in terms of the spectral density, and hence an approximate likelihood can be written down. Let

$$\zeta_n(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^{n} X_t e^{-it\lambda} \quad (-\pi < \lambda \leqslant \pi) \tag{1}$$

be the discrete Fourier transform of the data at frequency λ and let $J_n(\lambda) = |\zeta_n(\lambda)|^2$ be the periodogram. Then, for a fixed λ that is different from 0 or π, $J_n(\lambda)$ has a limiting exponential distribution with mean $f(\lambda)$; see Theorem 10.3.2 of Brockwell & Davis (1991).

Moreover, $\zeta(.)$ evaluated at two different frequencies, that are at least $2\pi/n$ apart, are asymptotically independent. To be more specific, let

$$F_n = \{\lambda_j = 2\pi j/n\colon\ -\pi < \lambda_j \leqslant \pi\} \tag{2}$$

be the 'Fourier frequencies'. If $\sum_{h=-\infty}^{\infty} |\gamma(h)|\,|h|^{\frac{1}{2}} < \infty$ and $E(X_t^4) < \infty$, then

$$\operatorname{cov}\{\zeta_n(\lambda_j), \zeta_n(\lambda_l)\} = O(n^{-1}) \tag{3}$$

for all $\lambda_j \neq \lambda_l$; see Brockwell & Davis (1991, p. 348). Hence, $\{J_n(\lambda_j)\colon \lambda_j > 0\}$ are approximately independent exponential random variables with means $f(\lambda_j)$, and an approximate likelihood of f may be written as

$$\mathscr{L}_n(f\,|\,X) = \exp\left[-\sum_{j=1}^{\lfloor n/2 \rfloor}\left\{\log f(\lambda_j) + \frac{J_n(\lambda_j)}{f(\lambda_j)}\right\}\right]. \tag{4}$$

Here and below $\lfloor y \rfloor$ denotes the greatest integer less than or equal to y. Negative frequencies are not included because $\zeta_n(-\lambda)$ is the complex conjugate of $\zeta_n(\lambda)$ and thus $J_n(-\lambda) = J_n(\lambda)$.

The Whittle likelihood is useful because it involves f directly, in contrast to the exact likelihood, which involves f indirectly through the autocovariances. Taniguchi (1987) and Taniguchi & Kakizawa (2000) provided results and references on higher-order efficiency of maximum likelihood estimators using (4) in autoregressive moving average models. Similar results for long memory time series were established by Fox & Taqqu (1986), Dahlhaus (1989) and Giraitis & Surgailis (1990) in parametric models and by Robinson (1994, 1995) in semiparametric models. In a nonparametric set-up, Whittle likelihood has also been used for automatic spline smoothing of the periodogram (Wahba, 1980), sieve maximum likelihood estimation (Chow & Grenander, 1985), penalised maximum likelihood estimation (Pawitan & O'Sullivan, 1994), polynomial spline fitting (Kooperberg et al., 1995), local-likelihood-based estimation (Fan & Kreutzberger, 1998) and Bayesian estimation of the spectral density (Carter & Kohn, 1997; Gangopadhyay et al., 1999), which is also discussed by N. Choudhuri, S. Ghosal and A. Roy in the unpublished report 'Bayesian estimation of the spectral density of a time series'.

Although one can use this simpler product likelihood for estimation, the properties of such estimators must be studied under the true distribution of the data. Unless a statistic has a fairly closed form, the study of its asymptotic properties is usually difficult. The product nature of the Whittle likelihood makes it relatively easy to study the asymptotic properties of a statistic under the distribution governing the Whittle likelihood, to be called the Whittle measure in the following. The goal of this paper is to show that the actual joint distribution of $\{J_n(\lambda_j)\colon j = 1, \ldots, \lfloor n/2 \rfloor\}$ for a Gaussian time series is mutually contiguous with the corresponding Whittle measure and thus asymptotic results obtained under the Whittle measure can be translated into those under the true distribution of the data.

The $O(n^{-1})$ rate of the covariances in equation (3) may lead one to conclude that the distribution of the triangular series $\{\zeta_n(\lambda_j)\colon \lambda_j \in F_n\}$ would be close to that of a vector of independent Gaussian random variables with increasing n. However, this is not always the case, as illustrated by the following example.

Example 1. Let $T_n = (T_{1,n}, \ldots, T_{n,n})$ have distribution $N(0, \Sigma_n)$ and let $S_n = (S_{1,n}, \ldots S_{n,n})$ have distribution $N(0, I_n)$, where I_n is the identity matrix of dimension n, the diagonal elements of Σ_n are all one and the off-diagonal elements are all $-(n-1)^{-1}$. Clearly $(\Sigma_n - I_n)$ converges uniformly to the zero matrix at rate n^{-1}. However, the distributions of T_n and S_n are quite different for each n; S_n has a full support on the entire $\mathbb{R}^n$ while T_n is confined to the $(n-1)$-dimensional subspace $1't=0$.

Thus, caution must be exercised while comparing the joint distributions of random vectors where the dimension increases to infinity along with n. There has been some investigation of the appropriateness of the Whittle likelihood as an approximation of the actual likelihood. Guyon (1982), Kunsch (1981) and Kent & Mohammadzadeh (1999) investigated the closeness of the two measures in the context of estimating spectra of Gaussian random fields, but the framework was mostly parametric. Strong approximation of the data ordinates by independent random variables that make up the Whittle likelihood was considered in Davis & Mikosch (1999) and Turkman & Walker (1990). However, these approximations say little about whether or not the asymptotic properties under the Whittle measure can be translated to those under the actual distribution and this is the motivation of this work. In § 2 we present our main result. Some applications are considered in § 3 and the proofs are presented in the Appendix.

2. Main result

For $j=1,\ldots,n-1$, define $r_j=\exp(i2\pi j/n)$ and $v_j=n^{-\frac{1}{2}}(r_j, r_j^2,\ldots,r_j^n)'$. For $j=1,\ldots,\lfloor n/2\rfloor$, define the cosine and sine vectors as $c_j=(v_j+v_{n-j})/\sqrt{2}$ and $s_j=i(v_j-v_{n-j})/\sqrt{2}$. Also define an n-vector $c_0=n^{-\frac{1}{2}}(1, 1, \ldots, 1)'$, an $n\times n$-orthogonal matrix

$$P_n=\begin{cases}(c_0, c_1, s_1, \ldots, 2^{-\frac{1}{2}}c_{n/2})', & \text{if } n \text{ is even},\\ (c_0, c_1, s_1, \ldots, c_{\lfloor n/2\rfloor}, s_{\lfloor n/2\rfloor})', & \text{if } n \text{ is odd},\end{cases} \tag{5}$$

and an $n\times n$ diagonal matrix

$$D_n=\begin{cases}2\pi\,\mathrm{diag}\{f(0), f(\lambda_1), f(\lambda_1), \ldots, f(\lambda_{\lfloor n/2\rfloor}), f(\lambda_{\lfloor n/2\rfloor})\}, & \text{if } n \text{ is odd},\\ 2\pi\,\mathrm{diag}\{f(0), f(\lambda_1), f(\lambda_1), \ldots, f(\lambda_{(n-2)/2}), f(\lambda_{(n-2)/2}), f(\lambda_{n/2})\}, & \text{if } n \text{ is even}.\end{cases} \tag{6}$$

Assume that the time series is Gaussian. Then $X\sim N(0,\Gamma_n)$, where $(\Gamma_n)_{ij}=\gamma(|i-j|)$. Let $P_{E,n}$ denote the true joint distribution of $Z=P_nX$, that is the joint distribution of $\{\zeta_n(\lambda_j)\colon \lambda_j\in F_n\}$, after a rearrangement of the components. Then $P_{E,n}=N(0, P_n\Gamma_nP_n')$, with density

$$(2\pi)^{-n/2}\det(\Gamma_n)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}z'(P_n\Gamma_nP_n')^{-1}z\}, \tag{7}$$

where det(.) denotes the determinant of a square matrix. Define the Whittle measure $P_{W,n}$ as the product measure of independent normals that give rise to the Whittle likelihood. Then $P_{W,n}=N(0, D_n)$, with density

$$(2\pi)^{-n/2}\det(D_n)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}z'D_n^{-1}z). \tag{8}$$

Since Z is a one-to-one function of the data X, $P_{E,n}$ and $P_{W,n}$ may also be regarded as the exact distribution and the Whittle approximate distribution of X. If the process has a nonzero mean then the term for the zero frequency is omitted. However, this makes only a minor change to our result.

Now we introduce the notion of contiguity for comparing the asymptotic behaviour of two sequences of probability measures that are defined on spaces that change with n. Contiguity plays vital roles in Le Cam's theory of efficiency of estimators and convergence of experiments.

Definition 1. *Let μ_n and ν_n be probability measures on some measurable space* $(\Omega_n, \mathscr{A}_n)$, *for* $n \geqslant 1$. *We say that μ_n and ν_n are mutually contiguous if, for every sequence of sets* $A_n \in \mathscr{A}_n$, $\mu_n(A_n) \to 0$ *if and only if* $\nu_n(A_n) \to 0$.

We make the following assumptions about the time series.

Assumption 1. We require that $\sum_{h=0}^{\infty} h^{\alpha}|\gamma(h)| < \infty$ for some $\alpha > 1$.

Assumption 2. We require that $f(\lambda) \neq 0$ for $\lambda \in [-\pi, \pi]$.

The following is the main result of this paper, followed by an obvious corollary that is used in most applications.

Theorem 1. *Let $\{X_t\}$ be a stationary Gaussian time series satisfying Assumptions* 1 *and* 2. *Then $P_{E,n}$ and $P_{W,n}$ are mutually contiguous.*

Corollary 1. *Under the assumptions of the theorem, the actual joint distribution of the periodogram ordinates* $\{J_n(\lambda_j)\colon j = 1, \ldots, \lfloor n/2 \rfloor\}$ *and the joint distribution of independent exponential random variables with means $f(\lambda_j)$, for $j = 1, \ldots, \lfloor n/2 \rfloor$, are mutually contiguous.*

3. Applications

The main usefulness of our result is that, regardless of the form of an estimator, if one can prove consistency or find the rate of convergence under the Whittle measure $P_{W,n}$, then the estimator is also consistent or, respectively, has the same rate of convergence, under the original Gaussian measure $P_{E,n}$. For instance, let $\hat{\theta}_n$ be an estimator of $\theta(f)$, a functional of the spectral density f, such that $\hat{\theta}_n - \theta(f) = O_p(\delta_n)$ under $P_{W,n}$, where δ_n is any sequence tending to zero as n tends to infinity. Suppose that, for any $M_n \to \infty$, the $P_{W,n}$ probability of the event $\{\|\hat{\theta}_n - \theta(f)\| > M_n\delta_n\}$ goes to 0 as $n \to \infty$. Then, by the mutual contiguity of $P_{W,n}$ and $P_{E,n}$, the $P_{E,n}$ probability of $\{\|\hat{\theta}_n - \theta(f)\| > M_n\delta_n\}$ also goes to zero, and hence $\hat{\theta}_n - \theta(f) = O_p(\delta_n)$ under $P_{E,n}$. A similar argument holds for posterior consistency or posterior rate of convergence in a Bayesian set-up. Also, by similar arguments, an in-probability asymptotic expansion, such as an Edgeworth expansion of a statistic or an asymptotic expansion of a posterior distribution as in Johnson (1970), obtained under the Whittle measure translates into a valid expansion under the actual distribution.

As a specific example of such applications of the contiguity result, we present a simpler derivation of the asymptotic results for the local-likelihood-based estimator of the log spectral density in Fan & Kreutzberger (1998). The problem can be written as a regression model,

$$Y_j = \log J_n(\lambda_j) = m(\lambda_j) + \varepsilon_j \quad (j = 1, \ldots, \lfloor n/2 \rfloor), \tag{9}$$

where the errors ε_j under $P_{W,n}$ are distributed as the logarithms of independent standard exponential random variables with density $\exp(-e^x + x)$ and $m(\lambda) = \log f(\lambda)$. Using the linear approximation $m(\lambda_j) \simeq m(\lambda) + m'(\lambda)(\lambda_j - \lambda)$ and localising the Whittle likelihood with a weight function $K_h(.)$, Fan & Kreutzberger (1998) estimated $m(\lambda)$ by $\hat{a}(\lambda)$, where

$(\hat{a}(\lambda), \hat{b}(\lambda))$ are obtained by minimising

$$\sum_{j=1}^{\lfloor n/2 \rfloor} [-\exp\{Y_j - a(\lambda) - b(\lambda)(\lambda_j - \lambda)\} + Y_j - a(\lambda) - b(\lambda)(\lambda_j - \lambda)]K_h(\lambda_j - \lambda) \tag{10}$$

for each λ. In order to derive the properties of the estimator, one needs to consider the complicated joint distribution of the ε_j's under $P_{E,n}$. However, the asymptotic results under $P_{W,n}$ are immediate from the results about local polynomial kernel regression under the independent exponential model of Fan et al. (1995). Then the derivation of the properties under $P_{E,n}$ is immediate if we apply our contiguity result.

Our result is particularly useful in the context of nonparametric Bayesian analysis where the posterior is obtained by updating the prior with the Whittle likelihood. The posterior distribution is analytically intractable, which renders any direct derivation of its asymptotic properties infeasible. Consistency and the rates of convergence of the posterior distribution have been well studied for independent and identically distributed observations. Extensions to independent nonidentically distributed observations have been recently addressed by Amewou-Atisso et al. (2003) and in the unpublished manuscript by N. Choudhuri, S. Ghosal and A. Roy 'Bayesian nonparametric binary regression'. However, similar results under dependence are not yet available and appear to be hard to obtain. Choudhuri et al., in the unpublished report mentioned in § 1, proposed a prior based on Dirichlet mixtures for the estimation of the spectral density. Under the measure $P_{W,n}$, they derived the consistency of the posterior distribution using an abstract theorem on posterior consistency for independent, nonidentically distributed observations. Using Theorem 1, they were able to show consistency under $P_{E,n}$. Indeed, one of the primary motivations of the present paper was to derive the posterior consistency of the nonparametric Bayesian procedure proposed in the report.

While the approach is elegant, abstract and powerful, one shortcoming is that the contiguity result applies only to a Gaussian time series. For the non-Gaussian case, Faÿ & Soulier (2001) proved a strong approximation result for independent and identically distributed observations. It will be of interest to extend the main result of this paper to general non-Gaussian processes.

Acknowledgement

We would like to thank the referees and the editor for their helpful comments and suggestions.

Appendix

Proof of Theorem 1

To prove Theorem 1, note that contiguity can be equivalently characterised by Le Cam's first lemma; see Theorem 1 of Le Cam & Yang (1990, p. 36) and Lemma 6.4 of van der Vaart (1998). If μ_n and ν_n are mutually absolutely continuous for all n, then mutual contiguity of μ_n and ν_n is equivalent to the tightness of $\Lambda_n = \log(d\nu_n/d\mu_n)$, the log likelihood ratio, under both μ_n and ν_n. Since $P_{E,n}$ and $P_{W,n}$ are mutually absolutely continuous for every n, it is enough to show that their log likelihood ratio,

$$\tfrac{1}{2}\{\log\det(\Gamma_n) - \log\det(D_n)\} + \tfrac{1}{2}\{Z'(P_n\Gamma_n P_n')^{-1}Z - Z'D_n^{-1}Z\}, \tag{A1}$$

is a tight, or stochastically bounded, sequence under both $P_{E,n}$ and $P_{W,n}$. This will be done by showing that the expectations and the variances of the log likelihood ratios in (A1) under both $P_{E,n}$ and $P_{W,n}$ are bounded sequences.

Let tr(.) denote the trace of a square matrix. Then, for a random vector $Z \sim N(0, \Sigma)$, we have

$$E(Z'AZ) = \text{tr}(A\Sigma), \quad \text{var}(Z'AZ) = 2\,\text{tr}(A\Sigma A\Sigma). \tag{A2}$$

Let $Q = Z'\{(P_n\Gamma_n P_n')^{-1} - D_n^{-1}\}Z$. Then, under $P_{E,n}$, $Z \sim N(0, P_n\Gamma_n P_n')$ and

$$E(Q) = \text{tr}(I_n - D_n^{-1}P_n\Gamma_n P_n'), \quad \text{var}(Q) = 2\,\text{tr}\{(I_n - D_n^{-1}P_n\Gamma_n P_n')^2\}. \tag{A3}$$

Under $P_{W,n}$, $Z \sim N(0, D_n)$ and

$$E(Q) = \text{tr}(P_n\Gamma_n^{-1}P_n'D_n - I_n), \quad \text{var}(Q) = 2\,\text{tr}\{(P_n\Gamma_n^{-1}P_n'D_n - I_n)^2\}. \tag{A4}$$

The theorem follows from the following two lemmas in view of (A1), (A3) and (A4).

LEMMA A1. *We have that* $\log\det(\Gamma_n) - \log\det(D_n) = O(1)$.

Proof. Note that, by Assumption 1, the first derivative of the spectral density is Lipschitz of order $\alpha - 1 > 0$; see Edwards (1982, p. 165). Applying the refinement of the distribution theorem in § 5.5 of Grenander & Szego (1984) with the function log, we have

$$\left|\log\det(\Gamma_n) - \frac{n+1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda)d\lambda\right| = O(1).$$

Now the lemma is proved by observing that

$$\frac{n+1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda)d\lambda - \log\det(D_n) = \frac{n+1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda)d\lambda - \sum_{\lambda_j \in F_n}\log f(\lambda_j)$$

is bounded as the Riemann sum converges at the rate n^{-1}. □

LEMMA A2. *The following sequences are bounded as* $n \to \infty$:

(i) $\text{tr}(I_n - D_n^{-1}P_n\Gamma_n P_n')$,
(ii) $\text{tr}\{(I_n - D_n^{-1}P_n\Gamma_n P_n')^2\}$,
(iii) $\text{tr}(P_n\Gamma_n^{-1}P_n'D_n - I_n)$,
(iv) $\text{tr}\{(P_n\Gamma_n^{-1}P_n'D_n - I_n)^2\}$.

Proof. (i) Note that $D_n = P_n\Gamma_n P_n' + H_n$, where the entries of H_n are uniformly $O(n^{-1})$ by Proposition 4.5.2 of Brockwell & Davis (1991). Now $\text{tr}(I_n - D_n^{-1}P_n\Gamma_n P_n') = \text{tr}(D_n^{-1}H_n)$. By Assumption 2 and the continuity of f, there exist positive constants m and M such that $m < f(\lambda) < M$ for $-\pi \leqslant \lambda \leqslant \pi$. Thus $1/M < 1/f < 1/m$ and the elements of D_n^{-1} are bounded by $1/m$, so that (i) follows.

(ii) With H_n as in (i), the expression in (ii) is $\text{tr}(D_n^{-1}H_n)^2 \leqslant m^{-2}\,\text{tr}(H_n^2)$. Clearly, the elements of H_n^2 are uniformly $O(n^{-1})$, and so (ii) follows.

(iii) Consider the variance covariance matrix $\tilde{\Gamma}_n$ of $(\tilde{X}_1, \ldots, \tilde{X}_n)$, where $\tilde{X}_t$ is a Gaussian time series with spectral density $1/f$. Since f is a continuous nonzero function on the compact interval $[-\pi, \pi]$, so is $1/f$. Thus, Assumptions 1 and 2 are clearly satisfied for the process with spectral density $1/f$. Also, define the $n \times n$ matrix

$$\tilde{D}_n = \begin{cases} 2\pi\,\text{diag}\left\{\dfrac{1}{f(0)}, \dfrac{1}{f(\lambda_1)}, \ldots, \dfrac{1}{f(\lambda_{\lfloor n/2 \rfloor})}, \dfrac{1}{f(\lambda_{\lfloor n/2 \rfloor})}\right\}, & \text{if } n \text{ is odd,} \\[2ex] 2\pi\,\text{diag}\left\{\dfrac{1}{f(0)}, \dfrac{1}{f(\lambda_1)}, \dfrac{1}{f(\lambda_1)}, \ldots, \dfrac{1}{f(\lambda_{(n-2)/2})}, \dfrac{1}{f(\lambda_{(n-2)/2})}, \dfrac{1}{f(\lambda_{n/2})}\right\}, & \text{if } n \text{ is even.} \end{cases}$$

Then, by Proposition 4.5.2 of Brockwell & Davis (1991),

$$\tilde{D}_n = P_n\tilde{\Gamma}_n P_n' + \tilde{H}_n,$$

where the elements of the matrix $\tilde{H}_n$ are also uniformly $O(n^{-1})$. Now the expression in (iii) can be written as

$$\operatorname{tr}(P_n \Gamma_n^{-1} P_n' H_n) = \operatorname{tr}(P_n \tilde{\Gamma}_n P_n' H_n) + \operatorname{tr}\{P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' H_n\}. \tag{A5}$$

The first term on the right-hand side of (A5) is

$$\operatorname{tr}(P_n \tilde{\Gamma}_n P_n' H_n) = \operatorname{tr}(\tilde{D}_n H_n) - \operatorname{tr}(\tilde{H}_n H_n).$$

Since $1/M < 1/f < 1/m$, the elements of $\tilde{D}_n$ are bounded. Since the elements of H_n are uniformly $O(n^{-1})$, it follows that $\operatorname{tr}(\tilde{D}_n H_n)$ is bounded. Furthermore, since the elements of $\tilde{H}_n$ are also uniformly $O(n^{-1})$, we see that $\operatorname{tr}(\tilde{H}_n H_n)$ is also bounded.

Now we shall bound the second term on the right-hand side of (A5). For a $n \times n$ matrix A_n define the ℓ_2 norm $|A_n|$ by the equation

$$|A_n| = \{\operatorname{tr}(A_n A_n')\}^{1/2} = \left(\sum_{l=1}^{n} \sum_{j=1}^{n} a_{lj}^2\right)^{\frac{1}{2}}$$

and the operator norm $\|A_n\|$ as $\|A_n\| = \sup\{\|A_n x\| : \|x\| = 1\}$, where x is an n-dimensional column vector. These two norms are related by $\|A_n\| \leqslant |A_n| \leqslant \sqrt{n}\|A_n\|$ and, for arbitrary square matrices A_n and B_n, $\operatorname{tr}(A_n B_n) \leqslant |A_n|\,|B_n|$ and $|A_n B_n| \leqslant \|A_n\|\,|B_n|$. Then, by (A1.1) in Dzhaparidze (1985),

$$|\operatorname{tr}\{P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' H_n\}| \leqslant |P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'|\,|H_n|.$$

Since the elements of H_n are uniformly $O(n^{-1})$, $|H_n| = O(1)$. It is therefore enough to show that $|P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'|$ is bounded. From (A1.2) in Dzhaparidze (1985),

$$\begin{aligned} |P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'| &= [\operatorname{tr}\{P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'\}]^{\frac{1}{2}} \\ &= [\operatorname{tr}\{\Gamma_n^{-1} - \tilde{\Gamma}_n)(\Gamma_n^{-1} - \tilde{\Gamma}_n)\}]^{\frac{1}{2}} = |\Gamma_n^{-1} - \tilde{\Gamma}_n| \\ &= |\Gamma_n^{-1}(I_n - \Gamma_n \tilde{\Gamma}_n)| \leqslant \|\Gamma_n^{-1}\|\,|I_n - \Gamma_n \tilde{\Gamma}_n|. \end{aligned} \tag{A6}$$

Now $\|\Gamma_n^{-1}\|$ is bounded by part (2) of Lemma A1.2 in Dzhaparidze (1985) and $|I_n - \Gamma_n \tilde{\Gamma}_n|$ is the square root of $\operatorname{tr}\{(I_n - \Gamma_n \tilde{\Gamma}_n)^2\}$, which is bounded by Lemma A1.4 of Dzhaparidze (1985). Thus (iii) follows.

(iv) The expression in (iv) can be written as

$$\begin{aligned} &\operatorname{tr}\{(P_n \Gamma_n^{-1} P_n' D_n - I_n)(P_n \Gamma_n^{-1} P_n' D_n - I_n)\} \\ &\quad = \operatorname{tr}\{\Gamma_n^{-1}(P_n' D_n P_n) \Gamma_n^{-1}(P_n' D_n P_n)\} - 2\operatorname{tr}\{\Gamma_n^{-1}(P_n' D_n P_n)\} + \operatorname{tr}(I_n) \\ &\quad = \operatorname{tr}\{\Gamma_n^{-1}(\Gamma_n + P_n' H_n P_n) \Gamma_n^{-1}(\Gamma_n + P_n' H_n P_n)\} - 2\operatorname{tr}\{\Gamma_n^{-1}(\Gamma_n + P_n' H_n P_n)\} + \operatorname{tr}(I_n) \\ &\quad = \operatorname{tr}(P_n \Gamma_n^{-1} P_n' H_n) + 2\operatorname{tr}(P_n \Gamma_n^{-1} P_n' H_n P_n \Gamma_n^{-1} P_n' H_n). \end{aligned}$$

The first term has been shown to be bounded in part (iii). Now

$$\{\operatorname{tr}(P_n \Gamma_n^{-1} P_n' H_n P_n \Gamma_n^{-1} P_n' H_n)\}^{\frac{1}{2}} = |P_n \Gamma_n^{-1} P_n' H_n|.$$

However,

$$|P_n \Gamma_n^{-1} P_n' H_n| \leqslant |P_n \tilde{\Gamma}_n P_n' H_n| + |P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' H_n|.$$

Since $|P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' H_n| \leqslant P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'|\,|H_n|$, $|H_n|$ is bounded and $|P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n'|$ has been shown to be bounded in part (iii), it follows that $|P_n(\Gamma_n^{-1} - \tilde{\Gamma}_n) P_n' H_n|$ is bounded. Also,

$$|P_n \tilde{\Gamma}_n P_n' H_n| = |(\tilde{D}_n - \tilde{H}_n) H_n| \leqslant |\tilde{D}_n H_n| + |\tilde{H}_n H_n|.$$

Since $\tilde{D}_n$ is diagonal with entries uniformly bounded, the entries of $\tilde{D}_n H_n$ are also uniformly bounded. By (A1.2) in Dzhaparidze (1985),

$$|\tilde{H}_n H_n| \leqslant \|\tilde{H}_n\| \, |H_n| \leqslant |\tilde{H}_n| \, |H_n|.$$

Thus both factors are bounded and (iv) follows. □

References

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K. & Ramamoorthi, R. V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli* **9**, 291–312.

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

Carter, C. K. & Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *J. R. Statist. Soc.* B **59**, 255–68.

Chow, Y. S. & Grenander, U. (1985). A sieve estimate for the spectral density. *Ann. Statist.* **13**, 998–1010.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–66.

Davis, R. A. & Mikosch, T. (1999). The maximum of the periodogram of a non-Gaussian sequence. *Ann. Prob.* **27**, 522–36.

Dzhaparidze, K. (1985). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. New York: Springer-Verlag.

Edwards, R. E. (1982). *Fourier Series. A Modern Introduction*. New York: Springer-Verlag.

Fan, J. & Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25**, 359–69.

Fan, J., Heckman, N. E. & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Am. Statist. Assoc.* **90**, 141–50.

Faÿ, G. & Soulier, P. (2001). The periodogram of an i.i.d sequence. *Stoch. Process Appl.* **92**, 313–43.

Fox, R. & Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517–32.

Gangopadhyay, A. K., Mallick, B. K. & Denison, D. G. T. (1999). Estimation of spectral density of a stationary time series via an asymptotic representation of the periodogram. *J. Statist. Plan. Infer.* **75**, 281–90.

Giraitis, L. & Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Prob. Theory Rel. Fields* **1**, 87–104.

Grenander, U. & Szego, G. (1984). *Toeplitz Forms and their Applications*. New York: Chelsea.

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

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851–64.

Kent, J. T. & Mohammadzadeh, M. (1999). Spectral approximation to the likelihood for an intrinsic Gaussian random field. *J. Mult. Anal.* **70**, 136–55.

Kooperberg, C., Stone, C. J. & Truong, Y. K. (1995). Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.* **16**, 389–401.

Kunsch, H. (1981). Thermodynamics and statistical analysis of Gaussian random fields. *Z. Wahr. verw. Geb.* **58**, 407–21.

Le Cam, L. & Yang, G. L. (1990). *Asymptotics in Statistics*. New York: Springer-Verlag.

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

Robinson, P. M. (1994). Semiparametric analysis of long-memory time series. *Ann. Statist.* **22**, 515–39.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23**, 1048–72.

Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *J. R. Statist. Soc.* B **49**, 315–25.

Taniguchi, M. & Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer Verlag.

Turkman, B. F. & Walker, A. M. (1990). A stability result for the periodogram. *Ann. Prob.* **18**, 1765–83.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc.* **75**, 122–32.

Whittle, P. (1957). Curve and periodogram smoothing. *J. R. Statist. Soc.* B **19**, 38–63.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bull. Int. Statist. Inst.* **39**, 105–29.

[*Received October* 2002. *Revised July* 2003]