The Annals of Statistics
1994, Vol. 22, No. 4, 2089–2114

ON THE BEHAVIOR OF A CAPON-TYPE SPECTRAL DENSITY ESTIMATOR[1]

BY EVANGELOS E. IOANNIDIS

University of Heidelberg

In this paper we propose a Capon-type estimator for the spectrum of a stationary time series. This estimator may be viewed as an alternative to classical periodogram-based estimators. Its advantage is that it copes with the "leakage effect" by using implicitly automatic adaptive windowing.

We show its asymptotic equivalence to a random variable which is a quadratic form in the observations, thus obtaining the asymptotic normality of the Capon estimator. We also study its asymptotic bias and variance.

1. Introduction. We consider the estimation of the spectral density f of a zero mean stationary stochastic process $\{X_i\}_{i \in Z}$, $X_i \in \mathbf{R}$, when $X_1, \dots, X_T$ have been observed. In this context estimators which are linear/ quadratic in the observations have been widely used and studied. On the other hand, there exist some nonlinear/nonquadratic estimators, which are often used in applied sciences; these are considered to have better small sample properties than classical periodogram-based estimators. One of them was introduced by Capon (1969) for estimating the wavenumber spectrum of a homogeneous random field. It became known to engineers and geophysicists as a "high resolution estimator" or "maximum likelihood method" [see, e.g., Chave, Douglas and Filloux (1991)]. Pisarenko (1972) considered a generalization of Capon's method in the context of estimating the continuous spectrum of a univariate time series.

The Capon estimator is defined as follows:

$$\hat{f}_{d,T}(\lambda) \equiv \hat{f}_{d,c,T}(\lambda) := \frac{d}{2\pi}\left(\bar{b}_\lambda^t \hat{\Gamma}_{d,c,T}^{-1} b_\lambda\right)^{-1}, \qquad \lambda \in [-\pi, \pi], \tag{1.1}$$

where $\hat{\Gamma}_{d,T} \equiv \hat{\Gamma}_{d,c,T}$ [defined in (1.2) below] is an estimator of the $d \times d$ covariance matrix Γ_d of $(X_1, \dots, X_d)^t$, $d < T/2$; $b_\lambda = \{\exp(i\lambda t)\}_{t=0,\dots,d-1} \in \mathbf{C}^d$ and finally both "d" and "c" are smoothing parameters (see also Remark 2 below). (For $b \in \mathbf{C}^d$ we denote by b^t the transpose and by $\bar{b}$ the component-wise conjugation.)

As covariance matrix estimator we use the following "segment" covariance matrix estimator:

$$\hat{\Gamma}_{d,T} = \hat{\Gamma}_{d,c,T} := \frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^t, \quad \text{where } N = N_{d,c,T} := \left[\frac{T-d}{c}\right] + 1 \tag{1.2}$$

Received March 1991; revised April 1994.

[1]This work was supported by the Deutsche Forschungsgemeinschaft. It was written under the guidance of Professor R. Dahlhaus and is part of the author's Ph.D. thesis.

AMS 1991 *subject classifications*. Primary 62M15; secondary 62M10, 62E20, 62G05.

Key words and phrases. Time series analysis, spectral estimator, high resolution estimator, leakage effect, Capon estimator, maximum likelihood method (MLM), covariance matrix estimation.

and

$$Y_i = Y_i^{d,c} := \left(X_{(i-1)c+1}, \ldots, X_{(i-1)c+d}\right)^t$$
$$\text{(c-displaced segments of length d of the data).}$$

Observe that for $c = 1$ we obtain almost fully overlapping segments and that for $c = d$ we obtain disjoint segments. Note further that $\widehat{\Gamma}_{d,c,T}$ is unbiased and not "Toeplitz." (Subsequently, we suppress in our notation the dependence of several quantities on n, d and c, and that on d and c on T, where this does not lead to confusion.)

An important advantage of the Capon estimator when compared to periodogram-based estimators is the following: when using periodogram-based estimators, the data should be tapered in order to reduce leakage, especially when the spectrum contains strong peaks [see, e.g., Bloomfield (1976), Sections 5.1–5.3]. Conversely, when using the Capon estimator, this seems not to be the case, if the covariance matrix estimator $\widehat{\Gamma}_{d,T}$ involved is unbiased. A heuristic argument for this is given by McDonough (1979). His conjecture is that the Capon estimator copes automatically with leakage by introducing implicitly automatic adaptive windowing. Thus the problem of choosing an optimal data taper is avoided.

Several interpretations of the Capon estimator were given by Marzetta (1983), Byrne and Fitzgerald (1984) and Burg (1972). The asymptotic distribution of the Capon estimator was studied by Capon and Goodman (1970), Pisarenko (1972) and Subba Rao and Gabr (1989). They assumed that $\{X_i\}_{i \in Z}$ is Gaussian and that independent segments Y_i, $i = 1, \ldots, N$, of $\{X_i\}_{i \in Z}$ have been observed (i.e., that $\widehat{\Gamma}_{d,T}$ is exactly Wishart distributed). The covariance matrix estimator used in these papers is

$$\widehat{\Gamma}_{d,T} := \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^t.$$

For the case where $X_1, \ldots, X_T$ have been observed they proposed that Y_i be taken as disjoint segments of the data.

For the Capon estimator $\widehat{f}_{d,T}(\lambda)$, as defined in (1.1) and (1.2), we prove in Sections 2 and 3 its asymptotic normality and study its bias and variance assuming $d, T \to \infty$. We show that the coefficient of the first-order expansion term of the variance is minimized for $c = 1$ (almost fully overlapping segments). In our proof we use a refinement of Pisarenko's expansion argument, the theory of orthogonal polynomials of Szegö (1959) and for the cumulant calculations the concept of "L^T functions" of Dahlhaus (1983). We replace the assumption that $\{X_i\}_{i \in Z}$ is Gaussian, made in the previously mentioned papers, by assumption (C) below. Furthermore, we drop their assumption that independent segments of $\{X_i\}_{i \in Z}$ have been observed; thus in the context where $X_1, \ldots, X_T$ have been observed, we can allow $d, T \to \infty$ simultaneously.

A technical result of independent interest is Lemma 2 in which $\|\widehat{\Gamma}_{d,c,T} - \Gamma_d\| \to_P 0$ is proved under the assumption that $cd^{1+\varepsilon}T^{-1} \to 0$. (If A is a real, symmetric matrix, then $\|A\| := \sup\{\|Ax\|_2, \|x\|_2 = 1\}$ denotes the operator norm of A.) This convergence is also used in the context of nonparametric spectral density estimation or prediction of time series via autoregressive approximation [e.g., Berk (1974), Shibata (1980) and Lewis and Reinsel (1985)]. In these papers $\|\widehat{\Gamma}_{d,c,T} - \Gamma_d\| \to_P 0$ is proved under the stronger assumption $d^2 T^{-1} \to 0$ and $c = 1$.

2. Asymptotic results. Let $c_u := E(X_t X_{t+u})$ denote the covariance. We assume that the spectral density f of the time series exists and is bounded above and away from 0. This assures that the $d \times d$ covariance matrix Γ_d [see (1.1) above] is positive definite for each d. We also make the general assumption that the distribution of $(X_1, \dots, X_T)$ on $\mathbf{R}^T$ is absolutely continuous with respect to the Lebesgue measure for each T. This also assures that $\widehat{\Gamma}_{d,T}$, as defined in (1.2), is positive definite with probability 1 if $d < N_{d,c,T}$. To confirm this, first observe that $\widehat{\Gamma}_{d,T}$ is always positive semidefinite. If it were singular and $d < N_{d,c,T}$, then the $d \times d$ matrix with columns $Y_1, \dots, Y_d$ would also be singular. In this case, for example, the last component of Y_d could be written as a function of the other elements of the matrix. The assumption of absolute continuity assures that this cannot happen with positive probability.

We introduce the following assumptions:

(A) f is continuous and $0 < m < f < M$. Further it fulfills

$$\exists\, C\colon |f(x) - f(y)| \le C\big|\log|x-y|\big|^{-1}\ \forall\, x, y \in [-\pi, \pi].$$

(B) The rth derivative of $1/f$ [denoted by ${}^{(r)}(f^{-1})$] is α-Lipschitz:

$${}^{(r)}\big(f^{-1}\big) \in \mathrm{Lip}_*^{\alpha} := \big\{g \mid \omega_*(g,\delta) = O(\delta^{\alpha}) \text{ as } \delta \to 0+\big\},$$

where

$$\omega_*(g,\delta) := \sup\big\{|g(x+h) + g(x-h) - 2g(x)|,\ x \in \mathbf{R},\ |h| \le \delta\big\}.$$

(C) $\{X_i\}_{i \in Z}$ has higher-order spectral densities $f^{(k)}$, which are absolutely bounded above for all orders k:

$$\big\|f^{(k)}\big\|_{\infty} < \infty \qquad \forall\, k > 0.$$

A spectral density of higher order is defined as a function $f^{(k)}\colon [-\pi, \pi]^{k-1} \to \mathbf{C}$, such that

$$\operatorname{cum}\big(X_{t_1}, \dots, X_{t_k}\big) = \int f^{(k)}(\alpha_1, \dots, \alpha_{k-1}) \times \exp\left(i \sum_{j=1}^{k-1} t_j \alpha_j - t_q i \sum_{j=1}^{k-1} \alpha_j\right) \mathbf{d}(\alpha_1, \dots, \alpha_{k-1}),$$

where $\operatorname{cum}(X_{t_1}, \dots, X_{t_k})$ denotes the cumulant of $(X_{t_1}, \dots, X_{t_k})$.

(Integrals are always taken over $[-\pi,\pi]^r$, for some r depending on the integration measure. Here $r = k-1$.)

We then have the following relations:

$$(2.1)\quad c_u = \int f(\lambda)\exp(i\lambda u)\,\mathbf{d}\lambda, \qquad \Gamma_d = \int f(\lambda) b_\lambda \overline{b}_\lambda^t\,\mathbf{d}\lambda, \qquad b_\lambda \text{ as in (1.1)},$$

and

$$f(\lambda) := \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c_u \exp(-i\lambda u).$$

For $\lambda \in \mathbf{R}$ we introduce the function

$$(2.2)\qquad \widetilde{f}_d(\lambda) := \frac{d}{2\pi}\left(\overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda\right)^{-1},$$

which is a theoretical quantity corresponding to the Capon estimator.

The main theorem of this paper follows. It states that the Capon estimator is an asymptotically unbiased, consistent estimator of f. Moreover, it states that $\sqrt{Td^{-1}}(\widehat{f}_{d,c,T} - f)$ is asymptotically normal [see statement (b) of the theorem].

For fixed $v \in \mathbf{R}^+$ and $\lambda_k \in [-\pi,\pi]$, $k = 1,\dots,K$, let $\zeta^v = (\zeta_1^v,\dots,\zeta_K^v) \in \mathbf{R}^K$ be a Gaussian random variable with expectation 0 and covariances given by $\operatorname{cov}(\zeta_j^v,\zeta_k^v) = v[\delta_{\lambda_j+\lambda_k} + \delta_{\lambda_j-\lambda_k}]$; here $\delta(\lambda)$ is defined as equal to 1 if $\lambda = 2k\pi$, $k \in \mathbf{Z}$; otherwise it is equal to 0. The asymptotic variance of the (standardized) estimator is given by $\lim_{T\to\infty}\theta(d_T/c_T)$, where

$$\theta(x) := x^{-1}\sum_{|u|<x}\left[1 - |u|x^{-1}\right]^2, \qquad x \in \mathbf{R}^+.$$

We therefore call θ the "variance function" (see Section A.3).

THEOREM 1. *Suppose that* (A), (B) *and* (C) *hold and that the sequences* c_T, d_T *fulfill*:

(i) $cd^{1+\varepsilon}/T \to 0$ *for some* $\varepsilon > 0$.

(ii) $c/d \le C$ *for some constant* $C < \infty$.

(iii) $d^{-\beta}\ln(T)(\ln(d))^2\ln(c) \to 0$ [*where* $\beta := (r+\alpha)/(1+r+\alpha)$, r, α *as in* (B)] *as* $T\to\infty$.

Then setting $v := \lim_{T\to\infty}\theta(d_T/c_T)$ (*assuming that the limit exists*) *we have*

(a)

$$\sqrt{Td^{-1}}\left\{\frac{\widehat{f}_{d,c,T}}{\widetilde{f}_d}(\lambda_k) - 1\right\}_{k=1,\dots,K} \Rightarrow_D \zeta^v.$$

(b) *If, in addition*, $Td^{-(1+2\gamma)}\ln^4(d) \to 0$, *where* $\gamma := (r+\alpha)\wedge 1$ [r, α *as in* (B)], *then*

$$\sqrt{Td^{-1}}\left\{\frac{\widehat{f}_{d,c,T}}{f}(\lambda_k) - 1\right\}_{k=1,\dots,K} \Rightarrow_D \zeta^v.$$

REMARK 1 (Choice of c). From the properties of θ (Lemma A.4) it is clear that the variance of the standardized estimator tends to the minimum possible value 2/3 when $c/d \to 0$, while it equals 1 when $c = d$ (disjoint segments). Moreover, for fixed d it is minimized by $c = 1$. The bound of $E\|\widehat{\Gamma}^*_{d,T} - I\|$ is also minimized by $c = 1$ (see Lemma 2 below). Note that for $c = d$ Theorem 1(i) becomes $d^{2+\varepsilon}/T \to 0$. We therefore propose to use $c = 1$.

REMARK 2 (Interpretation of d). The role of d in the Capon estimator or equivalently in $I^*_{d,T}$ is analogous to the inverse h^{-1} of the bandwidth "h" in the kernel-smoothed periodogram $I_T * K_h$, where $I^T(\lambda) := (2\pi T)^{-1}|\Sigma_{t=1}^T X_t e^{i\lambda t}|^2$, $K_h(x) := h^{-1}K(xh^{-1})$ and K is a kernel. This may be seen by the comparison of bias and variance: $\mathrm{var}(I^*_{d,1,T}) \approx T^{-1}\int(\Delta^d)^2 \approx \frac{2}{3}d/T$, where Δ^d is the Fejer kernel (see Section 3.3) and $\mathrm{var}(I^T * K_h/f) \approx (Th)^{-1}\int K^2$. In the same way if f fulfills (B) with $r = 0$, then from Lemma 1 below, it follows that (a) the bias of the Capon estimator is of order $O(d^{-\alpha}\ln(d))$ and (b) the bias of the autoregressive estimator of order p is $O(p^{-\alpha}\ln(p))$. On the other hand, the bias of the kernel-smoothed periodogram is, as can easily be seen, $O(h^\alpha)$ if K is Lipschitz continuous. If one imposes stronger regularity conditions on f, for example, that the kth derivative of f is absolutely bounded from above (it follows ${}^{(k-1)}f \in \mathrm{Lip}^1$), then from Lemma 1 below, it follows that (a) the bias of the Capon estimator is of an order not smaller than $O(d^{-1})$ and (b) the bias of the autoregressive estimator of order p is $O(p^{-k}\ln(p))$. Meanwhile, the bias of the kernel-smoothed periodogram becomes, as can easily be seen, $O(h^k)$ if the first $k-1$ moments of K vanish.

REMARK 3 (Choice of d). If c is equal to 1 (following Remark 1), the problem of choosing d arises. This may be done by using the Kullback–Leibler distance [Hurvich and Tsai (1989)], leading to an AIC-type criterion, which consists of taking $\widehat{d} := \mathrm{argmin}_{d \le d_{\max}} \int \ln(\widehat{f}_{d,1,T})(\lambda)\,\mathbf{d}\lambda + d/(T-d)$. This means that one computes $\int \ln(\widehat{f}_{d,1,T})(\lambda)\,\mathbf{d}\lambda$ for $d = 1, \dots, d_{\max}$ and uses the value of d which minimizes $\int \ln(\widehat{f}_{d,1,T})(\lambda)\,\mathbf{d}\lambda + d/(T-d)$ over d.

REMARK 4 (Efficiency considerations). In order to compare the performance of the Capon estimator with other estimators, we consider the bounds of the convergence rate to 0 of any error criterion, which has the form "squared bias + variance." Assuming $f \in \mathrm{Lip}^\alpha$, the error of the Capon estimator, the autoregressive estimator and the kernel-smoothed periodogram have almost the same rate [up to $\ln(T)$ terms]; for example, the error of the Capon estimator is obtained by the minimization of $d^{-2\alpha}\ln^2(T) + dT^{-1}$ (see Lemma 1), resulting in $O(T^{-(2\alpha)/(2\alpha+1)}\ln^2(T))$ (the power of the ln term has to be doubled when $\alpha = 1$). This is almost the same (up to the ln terms) as the rate for the kernel-smoothed periodogram [Parzen (1957)], and the one of the autoregressive estimator. On the other hand, if one imposes stronger regularity conditions on f, for example, that the kth derivative of f is absolutely bounded from above (it follows ${}^{(k-1)}f \in \mathrm{Lip}^1$), the error of the Capon estimator has a slower rate than those of the kernel-smoothed periodogram and the autoregressive estimator. The error of the Capon estimator remains, in general, of order not smaller than $O(T^{-2/3})$. The error of the kernel-smoothed periodogram becomes $O(T^{-(2k)/(2k+1)})$ if the first $k-1$ moments of K vanish; already for $k = 2$ and K symmetric this is $O(T^{-4/5})$. The error of the autoregressive estimator is of order $O(T^{-(2k)/(2k+1)} \ln^2(T))$ (which follows easily from Lemma 1). The difference between the Capon estimator, on the one hand, and the preceding two, on the other, is due to the different order of the bias (see Remark 2 above and Lemma 1 below). Note also that $O(T^{-(2k)/(2k+1)})$ is the optimal rate under the conditions used here [see, e.g., Rosenblatt (1985), Section 5.6].

REMARK 5 (The nonzero mean case). In the case where the process $\{X_i\}_{i \in Z}$ has an unknown mean μ, one may use the Capon estimator based on the centered data $X_i - \overline{X}_T$, where $\overline{X}_T$ is the empirical mean. Then our conjecture is that *Theorem* 1 *will also hold for the resulting estimator*. We indicate the asymptotics in Section A.4.

The proof of Theorem 1 is obtained from several technical lemmas in Section 3. We now sketch the basic idea of the proof and present two results that are of independent interest. The first concerns the bias of the Capon estimator. It gives upper bounds for the convergence rate of $\tilde{f}_d - f$ to 0. By $\{\phi_k(\lambda)\}_{k \in N}$ we denote the system of orthogonal polynomials associated with f (see Section A.2). The quantity $\tilde{f}_d^{-1}$ may be written as the mean over the squared modulus of orthogonal polynomials [see (A.2.2)]:

$$\tilde{f}_d^{-1} = d^{-1} \sum_{i=0}^{d-1} |\phi_i(\lambda)|^2.$$

Since lower-order polynomials are also involved in this mean, it is clear that in general $\tilde{f}_d - f$ will be of an order not smaller than d^{-1}.

LEMMA 1. *Assuming that* (A) *and* (B) *hold and that r and α are as in* (B), *the following hold*:

(a) $$\big\| |\phi_d|^2 - f^{-1} \big\|_\infty = O\big(d^{-(r+\alpha)} \ln(d)\big).$$

(b) $$\big\| \tilde{f}_d^{-1} - f^{-1} \big\|_\infty = \begin{cases} O\big(d^{-[(r+\alpha)\wedge 1]} \ln(d)\big), & \text{if } r+\alpha \neq 1, \\ O\big(d^{-1} \ln^2(d)\big), & \text{if } r+\alpha = 1. \end{cases}$$

(The bounds of the form $O[a_T \ln(b_T)]$ with $a_T, b_T \in \mathbf{R}$ are always to be read as $O[a_T]$ when $b_T \equiv 1$.)

To obtain the asymptotic distribution of the Capon estimator, we need an expansion of it. It turns out that it is technically more convenient to expand the standardized quantity

$$\widehat{f}_{d,c,T}/\widetilde{f}_d = \left(\overline{b}_\lambda^{*t}(\widehat{\Gamma}^*_{d,T})^{-1}b^*_\lambda\right)^{-1},$$

where

$$\widehat{\Gamma}^*_{d,T} \equiv \widehat{\Gamma}^*_{d,c,T} := U_d^{-1}\widehat{\Gamma}_{d,c,T}\left(U_d^t\right)^{-1} \quad \text{and} \quad b^*_\lambda := U_d^{-1}b_\lambda/\|U_d^{-1}b_\lambda\|_2.$$

Here $\Gamma_d = U_d U_d^t$ is the Cholesky decomposition of Γ_d (i.e., $U_d = U_{d,\Gamma}$ is a lower triangular $d \times d$ matrix). Quantities involving a standardization with U_d^{-1} will be denoted by $*$. Thus $\widehat{\Gamma}^*_{d,T}$ is a "standardization" of $\widehat{\Gamma}_{d,T}$.

By a Neumann expansion of $(\widehat{\Gamma}^*_{d,T})^{-1}$ [e.g., in Yosida (1980). Section 2.1, Theorem 2], we show that $\widehat{f}_{d,c,T}/\widetilde{f}_d$ is asymptotically equivalent to

$$I^*_{d,T}(\lambda) \equiv I^*_{d,c,T}(\lambda) := \overline{b}_\lambda^{*t}\widehat{\Gamma}^*_{d,T}b^*_\lambda.$$

$I^*_{d,c,T}$ is a random variable which involves the (unknown) true covariance matrix but is quadratic in the observations and can therefore be studied by standard cumulant methods. For the expansion to be valid we need $\|\widehat{\Gamma}^*_{d,T} - I\| \to_P 0$, where I is the $d \times d$ identity matrix. Conditions under which this holds are of more general interest (see the Introduction).

LEMMA 2. *Assuming that* (A) *and* (C) *hold and that the sequences* c_T, d_T *fulfill* $dc\ln(T)\ln(c)/T \to 0$ *and* $c/d < C$ *(for some* C*), we have*

$$\|\widehat{\Gamma}^*_{d,T} - I\| = O_P\left(\sqrt{cd^{1+\varepsilon}\ln(T)\ln(c)\ln(d)/T}\right), \qquad \varepsilon > 0 \textit{ arbitrary}.$$

3. Detailed results—proofs.

3.1. *Bias considerations.* In this section we prove Lemma 1 by utilizing properties of orthogonal polynomials, which are stated in the Appendix.

PROOF OF LEMMA 1. (a) The result follows directly from Lemma A.3(a), taking $p_n \equiv n \equiv d$. Observe that by (A.2.4) we have $|\psi_{d,d}|^2 \equiv |t_d|^2$.

Part (b) follows directly from (a), since $\|\widetilde{f}_d^{-1} - f^{-1}\|_\infty \le d^{-1}\Sigma_{p=0}^{d-1}\||\phi_p|^2 - f^{-1}\|_\infty$. $\square$

3.2. *Asymptotic expansion.* In this section we prove the asymptotic equivalence between $\widehat{f}_{d,c,T}/\widetilde{f}_d$ and $I^*_{d,c,T}$.

LEMMA 3. *Assuming that* $\sqrt{T/d}\|\widehat{\Gamma}^*_{d,T} - I\|^2 \to_P 0$, $T \to \infty$, *then* $\sqrt{T/d}[\widehat{f}_{d,T}/\widetilde{f}_d - I^*_{d,T}] \to_P 0$.

PROOF. First note that

$$\frac{\widehat{f}_{d,c,T}}{\widetilde{f}_d} = \frac{\left(\overline{b}_\lambda^t (U_d^t)^{-1} U_d^t \widehat{\Gamma}_{d,T}^{-1} U_d U_d^{-1} b_\lambda\right)^{-1}}{\left(\overline{b}_\lambda^t (U_d^t)^{-1} U_d^{-1} b_\lambda\right)^{-1}} = \left(\overline{b}_\lambda^{*t} (\widehat{\Gamma}_{d,T}^*)^{-1} b_\lambda^*\right)^{-1}.$$

Now let the following event be denoted by $A_N = \{\|\widehat{\Gamma}_{d,T}^* - I\| < 1/2\}$. Then on A_N we may expand $(\widehat{\Gamma}_{d,T}^*)^{-1}$:

$$(\widehat{\Gamma}_{d,T}^*)^{-1} = \left[I - (I - \widehat{\Gamma}_{d,T}^*)\right]^{-1} = \sum_{j=0}^{\infty} (I - \widehat{\Gamma}_{d,T}^*)^j.$$

This implies that

$$\left(\overline{b}_\lambda^{*t} (\widehat{\Gamma}_{d,t}^*)^{-1} b_\lambda^*\right)^{-1} - 1 = \overline{b}_\lambda^{*t} (\widehat{\Gamma}_{d,T}^* - I) b_\lambda^* + O\left(\|\widehat{\Gamma}_{d,T}^* - I\|\right)^2 \quad \text{on } A_N.$$

Since $\overline{b}_\lambda^{*t} (\widehat{\Gamma}_{d,T}^* - I)\, b_\lambda^* = I_T^* - 1$, $\sqrt{T/d}\|\widehat{\Gamma}_{d,T}^* - I\|^2 \to_P 0$ and $P(A_N^C) \to 0$, the proof is complete. □

3.3. *Consistency.* In this section we prove Lemma 2 as a direct consequence of Lemma 4, stated below. Let $N = N_{d,c,T}$ as in (1.2).

LEMMA 4. *If* (A) *and* (C) *hold*, $K \in \mathbf{N}$ *is even*, $dN^{-1}\ln(N)\ln(c) \to 0$ *and* $c/d < C$ *(for some constant* C*), then we have*

$$E\,\mathrm{tr}[\widehat{\Gamma}_{d,T}^* - I]^K = O\left(N^{-K/2} d^{K/2+1} \left[\ln(N)\ln(c)\ln(d)\right]^{K/2}\right).$$

To prove this we introduce the following notation and state a further proposition.

We consider a $2 \times k$ table of variables which has the following form:

$$\begin{array}{cc} \alpha_1 & \beta_1 \\ \vdots & \vdots \\ \alpha_k & \beta_k \end{array}$$

and partitions $\mathcal{P}^{(k)} = \{P_1, \dots, P_S\}$ of the $2 \times k$ table, $s_i := |P_i|$. For a partition subset $P_i := (\kappa_1, \dots, \kappa_{s_i})$ (κ stands for some α and β), we denote $\widetilde{\kappa}_i := (\kappa_1, \dots, \kappa_{s_i - 1})$ and set $\kappa_{s_i} := -\Sigma_{j=1}^{s_i - 1} \kappa_j$. Let $\Sigma_{\mathrm{ip},(k)}$ denote summation over the indecomposable partitions of the $2 \times k$ table, $\Sigma_{\mathrm{ap},(k)}$ denote summation over *all* partitions of the $2 \times k$ table and $\Sigma_{\mathrm{ap}\setminus *,(k)}$ denote summation over *all* partitions of the $2 \times k$ table *excluding* those which contain a partition subset consisting of exactly one row of the table.

In this sequel let $\mathbf{b}_\lambda = \{e^{i\lambda t}\}_{t=0,\dots,T-1} \in \mathbf{C}^T$ and $\mathbf{X} = (X_1, \dots, X_T)^t$.

PROPOSITION 1. *Assuming that* (C) *holds, then for arbitrary* $T \times T$ *matrices* $A_j, j = 1, \dots, k$ *(using the notation in the* $2 \times k$ *table, e.g.,* S *is the number of partition subsets of* $\mathcal{P}^{(k)}, s_i := |P_i|$*) we have*

(a) $$\operatorname{cum}\left(\prod_{j=1}^{k} \mathbf{X}^t A_j \mathbf{X}\right) = \sum_{\mathrm{ap},(k)} \int_{[-\pi,\pi]^{2k-S}} \prod_{i=1}^{S} f^{(s_i)}(\widetilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}^t_{\alpha_j} A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i.$$

(b) $$\begin{aligned}&\operatorname{cum}(\mathbf{X}^t A_1 \mathbf{X}, \dots, \mathbf{X}^t A_k \mathbf{X})\\ &\quad = \sum_{\mathrm{ip},(k)} \int_{[-\pi,\pi]^{2k-S}} \prod_{i=1}^{S} f^{(s_i)}(\widetilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}^t_{\alpha_j} A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i.\end{aligned}$$

(c) $$\begin{aligned}&\operatorname{cum}\left(\prod_{j=1}^{k} [\mathbf{X}^t A_j \mathbf{X} - E\mathbf{X}^t A_j \mathbf{X}]\right)\\ &\quad = \sum_{\mathrm{ap}\backslash *,(k)} \int_{[-\pi,\pi]^{2k-S}} \prod_{i=1}^{S} f^{(s_i)}(\widetilde{\kappa}_i) \prod_{j=1}^{k} \mathbf{b}^t_{\alpha_j} A_j \mathbf{b}_{\beta_j} \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i.\end{aligned}$$

PROOF. The results follow from the product theorem for cumulants [Brillinger (1975), Theorem 2.3.2] by a straightforward calculation and by using the spectral representation of $\operatorname{cum}(X_1, \dots, X_q)$. □

Further let $E_j = E_j^{d,c,T} := [\mathbf{0}_{d \times (j-1)c} \mathbf{I}_d \, \mathbf{0}_{d \times (T-d-(j-1)c)}] \in \mathbf{R}^{d \times T}$. Then we have $Y_j = Y_j^{d,c} = E_j \mathbf{X}$. Denote the Fejer kernel by

$$\Delta^N(\lambda) := N^{-1} \left|\Theta^N(\lambda)\right|^2 \quad \text{where } \Theta^N(\lambda) := \sum_{t=1}^{N} e^{i\lambda t}$$

and also

$$K_d(\alpha, \beta) := \overline{b}^t_\alpha \Gamma_d^{-1} b_\beta.$$

For a given partition $\mathcal{P}^{(k)}$ of the $2 \times k$ table above, we denote

$$\mathbf{V}(\mathcal{P}^{(k)}) := N^{-k} \int_{[-\pi,\pi]^{2k-S}} \prod_{i=1}^{S} f^{(s_i)}(\widetilde{\kappa}_i) \prod_{i=1}^{k} \Theta^N(c\alpha_i + c\beta_i) \, K_d(-\alpha_i, \beta_{i+1}) \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i$$

and

$$\widetilde{\mathbf{V}}(\mathcal{P}^{(k)}) := N^{-k} \int_{[-\pi,\pi]^{2k-S}} \prod_{i=1}^{k} L^N(c\alpha_i + c\beta_i) \, L^d(\alpha_i + \beta_{i+1}) \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i,$$

where the L^N functions are defined in Section A.1. In these expressions indices are always taken mod(k), for example, $\beta_{k+1} \equiv \beta_1$.

PROOF OF LEMMA 4. The proof is given in two steps.

Step 1. We first prove

$$E\,\mathrm{tr}[\widehat{\Gamma}^*_{d,T} - I]^K = \sum_{\mathrm{ap}\backslash *, (K)} \mathbf{V}(\mathcal{P}^{(K)}). \tag{3.1}$$

From $I = (2\pi)^{-1} \int b_\lambda \overline{b}^t_\lambda \, \mathbf{d}\lambda$ (I is the $d \times d$ identity matrix), we obtain

$$\mathrm{tr}[\widehat{\Gamma}^*_{d,T} - I]^K = (2\pi)^{-K} \int \prod_{i=1}^{K} \overline{b}^t_{\lambda_i} [\widehat{\Gamma}^*_{d,T} - I] b_{\lambda_{i+1}} \prod_{i=1}^{K} \mathbf{d}\lambda_i.$$

Observe that $\overline{b}^t_{\lambda_i} \; [\widehat{\Gamma}^*_{d,T} - I] b_{\lambda_{i+1}} = \mathbf{X}^t A_i \mathbf{X} - \overline{b}^t_{\lambda_i} b_{\lambda_{i+1}}$, where $A_i := N^{-1} \times \Sigma_{j=1}^N E^t_j (U^t_d)^{-1} b_{\lambda_{i+1}} \overline{b}^t_{\lambda_i} U_d^{-1} E_j$. We have $E_j \mathbf{b}_\beta = \exp[-i(j-1)\,c\beta] b_\beta$ which yields

$$\mathbf{b}^t_\alpha A_j \mathbf{b}_\beta = N^{-1} \Theta^N (c\alpha + c\beta) b^t_\alpha \left(U^t_d\right)^{-1} b_{\lambda_{j+1}} \overline{b}^t_{\lambda_j} U_d^{-1} b_\beta \exp\left[-i(c\alpha + c\beta)\right].$$

With these we get directly from Proposition 1(c) that the expectation of $\mathrm{tr}[\widehat{\Gamma}^*_{d,T} - I]^K$ equals (using the notation of the $2 \times k$ table given previously, with $k = K$)

$$(2\pi)^{-K} N^{-K} \sum_{\mathrm{ap}\backslash *, (K)} \int \prod_{i=1}^{S} f^{(s_i)}(\widetilde{\kappa}_i) \prod_{i=1}^{K} \times \left[\Theta^N (c\alpha_i + c\beta_i) b^t_{\alpha_i} \left(U^t_d\right)^{-1} b_{\lambda_{i+1}} \overline{b}^t_{\lambda_i} U_d^{-1} b_{\beta_i}\right] \prod_{i=1}^{K} \mathbf{d}\lambda_i \prod_{i=1}^{S} \mathbf{d}\widetilde{\kappa}_i.$$

From this (3.1) follows after integration with respect to $\lambda_i, i = 1, \ldots, K$, noting that

$$(2\pi)^{-1} \int b^t_{\alpha_i} \left(U^t_d\right)^{-1} b_{\lambda_{i+1}} \overline{b}^t_{\lambda_{i+1}} U_d^{-1} b_{\beta_{i+1}} \, \mathbf{d}\lambda_{i+1} = b^t_{\alpha_i} \Gamma_d^{-1} b_{\beta_{i+1}} = K_d(-\alpha_i, \beta_{i+1}).$$

Step 2. We prove that

$$\widetilde{\mathbf{V}}(\mathcal{P}^{(K)}) = O\left(N^{-K/2} d^{K/2+1} \left[\ln(N) \ln(c) \ln(d)\right]^{K/2}\right), \tag{3.2}$$

where $\mathcal{P}^{(K)}$ is a fixed partition of the $2 \times K$ table not containing a one-row partition subset. This is sufficient for the assertion of the lemma because of (3.1) and since $|\mathbf{V}(\mathcal{P}^{(K)})| = O(\widetilde{\mathbf{V}}(\mathcal{P}^{(K)}))$ (because of Lemma A.2).

For this fixed partition $\mathcal{P}^{(K)}$, we will call *indecomposable row subtable* a union of rows of the original $2 \times K$ table, if and only if it can be written as a union of some partition subsets and cannot be split up any more in this sense. Similarly we will call *indecomposable diagonal subtable* a union of diagonals $\{\alpha_i, \beta_{i+1}\}$ of the original $2 \times K$ table, if and only if it can be written as a union of some partition subsets and cannot be split up any more in this sense.

Assuming that we can split up $\mathcal{P}^{(K)}$ into M row subtables, with $m_1, \dots, m_M$ rows, and into L diagonal subtables and further enumerate the elements of each partition subset P, in such a way that its last element appears in the first row of those containing elements of P, we will prove that

$$\widetilde{\mathbf{V}}(\mathcal{P}^{(K)}) = O\big(N^{M-K} d^{1+K-M} c^{M-K+L-1} \ln(N)^{K-M} \ln(c)^{K-M-L+1} \ln(d)^{M}\big). \tag{3.3}$$

This is sufficient for (3.2), since

(a) $M \le K/2$, as we assumed $m_l \neq 1 \,\forall\, l$;

(b) $dN^{-1}\ln(N)\ln(c) \to 0$ by assumption of the lemma.

(c) $M - K + L - 1 \le 0$, since each m row subtable contains maximally $m-1$ diagonal subtables.

To prove (3.3), we integrate with respect to each of the variables under the integral in $\widetilde{\mathbf{V}}(\mathcal{P}^{(K)})$ using Lemmas A.1(i) and A.1(ii). Note that this is possible since we have the structure needed to use them: each variable appears once with positive sign and once with negative sign in the arguments of the L^N factors as well as in the arguments of the L^d factors (if it appears at all). We use the following scheme: first we integrate with respect to the variables (which are not the last of their partition subsets) in the first row of the $2 \times K$ table, then in the second, and so on. We use Lemma A.1(i) whenever it is possible.

Observe that the factor $L^N(c\alpha_j + c\beta_j)$ appears in the integrand in $\widetilde{\mathbf{V}}$ as long as it has not been integrated with respect to either α_j or β_j. In this case we call the row j "unconnected." We also call "connecting" an unconnected row when we integrate with respect to α_j or β_j. In this process $L^N(0) = N$ will appear exactly when all rows of a row subtable have been connected; thus we obtain the factor N^M for the final bound. In the same way we also obtain the factor d^L. Finally, Lemma A.1(ii) will only—but not necessarily—be used when connecting a row, that is at most $K - M$ times. Moreover we claim that

$$\text{the number of times Lemma A.1(ii) will have to be used instead of Lemma A.1(i) does not exceed } K - M - L + 1. \tag{3.4}$$

Accordingly, in the final bound, the factor:

- $d\ln(N)\ln(c)/c$ will appear whenever Lemma A.1(ii) is used, thus $K - M - L + 1$ times;
- $\ln(N)$ will additionally appear whenever Lemma A.1(i) is used (on two L^N functions), thus $K - M$ times in total;
- $\ln(d)$ will appear whenever Lemma A.1(i) is used (on two L^d functions), thus $K - L + 1-$ [number of times Lemma A.1(ii) is used].

These yield

$$\widetilde{\mathbf{V}}(\mathcal{P}^{(K)}) < O(1)N^{M-K}d^{L}\left(\frac{d}{c}\right)^{K-M-L+1}\ln(N)^{K-M}\ln(c)^{K-M-L+1} \times \ln(d)^{(K-L+1)-(K-M-L+1)}.$$

This proves (3.3). It remains to prove (3.4). In view of our remark preceding (3.4), it is sufficient to prove (3.5).

(3.5) For each diagonal subtable, up to the one containing α_1 and β_k, one may use Lemma A.1(i) instead of Lemma A.1(ii) for connecting its last row.

PROOF OF (3.5): Let β_{j+1} be the element of the diagonal subtable considered in (3.5) with maximal index. Due to our enumeration we may assume—without loss of generality—that:

(i) β_{j+1} is not the last element of any partition subset.
(ii) The row $j+1$ has not yet been connected.
(iii) The row j has been connected and furthermore integration with respect to α_j, β_j has taken place, if they are not the last elements of a partition subset.

There are two cases:

(i) α_j is the last element of a partition subset P. Then P must contain β_{j+1}. The reason is that since β_{j+1} has maximal index in its diagonal subtable, our enumeration allows only β_j, β_{j+1} as probable elements of P. The assumption that there is no one-row partition subset excludes the case $\beta_{j+1} \notin P$.
(ii) (i) is not fulfilled; that is, α_j is not the last element of any partition subset P.

In both cases β_{j+1} does not occur in the argument of an L^d factor and thus the "$j+1$" row may be connected by using Lemma A.1(i). This proves (3.5). □

Finally, we prove Lemma 2.

PROOF OF LEMMA 2. Let $\alpha_{T,c,d} := cd^{1+\varepsilon}T^{-1}\ln(T)\ln(c)\ln(d)$. For any $M > 0$ and $K \in Z_+$, Chebyshev's inequality yields

$$P\Big[\sqrt{\alpha_{T,c,d}^{-1}}\|\widehat{\Gamma}^*_{d,c,T} - I\| > M\Big] \le \alpha_{T,c,d}^{-K} M^{-2K} E\,\mathrm{tr}\big[\widehat{\Gamma}^*_{d,T} - I\big]^{2K} \le M^{-2K} C_K d^{1-K\varepsilon}$$

for some constant C_K, according to Lemma 4. Choosing $K > (\varepsilon)^{-1}$ completes the proof, since $d^{1-K\varepsilon}$ will be bounded. □

3.4. *Cumulants of $I^*_{d,c,T}(\lambda)$.* In this section we study the second- and higher-order cumulants of the random variable $I^*_{d,T}(\lambda)$. Let δ denote the Dirac function, extended to be 2π-periodic, θ the "variance function," defined in Section 2, and $N = N_{d,c,T}$. We prove the following results.

LEMMA 5. *Under* (A), (B) *and* (C), *assuming* $T \to \infty$ *and* $c/d < C$ (*for some* C), *we have*

$$E[I^*_{d,c,T}(\lambda) - 1] = 0$$

and

$$\operatorname{cov}\left[\sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda), \sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\mu)\right] = \begin{cases} O(R_2), & \text{if } \lambda \not\equiv \pm\mu \bmod(2\pi), \\ \theta(d/c)\,[\delta_{\lambda+\mu} + \delta_{\lambda-\mu}] + O(R_1), & \text{otherwise}, \end{cases}$$

with

$$R_2 := d^{-2}\left[|\lambda - \mu|^{-2} + |\lambda + \mu|^{-2}\right] \ln(N)\ln(d)\ln(c) + R_1$$

and

$$R_1 := d^{-\beta}\ln(N)\ln(d)\ln(c) + d(cN)^{-1} \quad \text{where } \beta := \frac{r+\alpha}{1+r+\alpha}, r, \alpha \text{ as in (B)}.$$

LEMMA 6. *Under* (A), (B) *and* (C), *assuming* $T \to \infty$ *and* $c/d < C$ (*for some* C), *the following holds*:

$$\operatorname{cum}\left[\sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda_1), \ldots, \sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda_r)\right] = O\Big(\left[dc^{-1}N^{-1}\right]^{r/2-1}\left[\ln(N)\ln(c)\right]^{r-1}\ln(d)\Big).$$

In order to prove these two lemmas, we need the following result (using the same notation as in the $2 \times k$ table).

COROLLARY 1. *Assuming* (C), *the following holds*:

$$\begin{aligned} &\operatorname{cum}\left[\sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda_1), \ldots, \sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda_r)\right] \\ &\quad = d^{-3r/2}N^{-r/2}c^{r/2}(2\pi)^r \prod_{j=1}^{r} \tilde{f}_d(\lambda_j) \\ &\qquad \times \sum_{\mathrm{ip},(k)} \int \prod_{i=1}^{S} f^{(s_i)}(\tilde{\kappa}_i) \prod_{j=1}^{r} \left[K_d(\lambda_j, \alpha_j) K_d(-\lambda_j, \beta_j) \Theta^N(c\alpha_j + c\beta_j)\right] \prod_{i=1}^{S} \mathbf{d}\tilde{\kappa}_i. \end{aligned}$$

PROOF. The result is a direct consequence of Proposition 1. This is seen by observing that

$$I^*_{d,c,T}(\lambda) = 2\pi d^{-1}\tilde{f}_d(\lambda)\mathbf{X}^t A_\lambda \mathbf{X} \quad \text{with } A_\lambda := N^{-1}\textstyle\sum_{j=1}^{N} E_j^t \Gamma_d^{-1} \bar{b}_\lambda b_\lambda^t \Gamma_d^{-1} E_j$$

and that

$$\mathbf{b}_\alpha^t A_\lambda \mathbf{b}_\beta = K_d(\lambda_j, \alpha_j)K_d(-\lambda_j, \beta_j)N^{-1}\Theta^N(c\alpha_j + c\beta_j)\exp(-c\alpha_j - c\beta_j). \qquad \square$$

Using this corollary, one may prove the lemma concerning the first- and second-order cumulants of $I^*_{d,c,T}(\lambda)$.

PROOF OF LEMMA 5. The first statement concerning the expectation of $I^*_{d,c,T}(\lambda)$ follows by observing that $E\widehat{\Gamma}^*_{d,c,T} = I_d$.

Next we study the covariance structure of $I^*_{d,c,T}(\lambda)$. The indecomposable partitions of the 2×2 table are (i) $\{\alpha_1, \beta_2\}, \{\beta_1, \alpha_2\}$; (ii) $\{\alpha_1, \alpha_2\}, \{\beta_1, \beta_2\}$; and (iii) $\{\alpha_1, \alpha_2, \beta_1, \beta_2\}$. Thus by Corollary 1 and setting

$$\begin{aligned} P_{d,c,T}(f, \lambda, \mu) := \int & f(\alpha)f(\beta)\Delta^N(c\beta - c\alpha)K_d(\lambda, \alpha)K_d(-\mu, -\alpha) \\ & \times K_d(\mu, \beta)K_d(-\lambda, -\beta)\, \mathbf{d}\alpha\, \mathbf{d}\beta, \end{aligned}$$

we obtain

$$\begin{aligned} &\operatorname{cov}\left[\sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\lambda), \sqrt{\frac{Nc}{d}} I^*_{d,c,T}(\mu)\right] \\ &\quad = d^{-3}c(2\pi)^2 \widetilde{f}_d(\lambda)\widetilde{f}_d(\mu)\Big[P_{d,c,T}(f, \lambda, \mu) + P_{d,c,T}(f, \lambda, -\mu) \\ &\qquad + \int f^{(4)}(\alpha, \beta, \gamma)\Delta^N(c\alpha + c\beta)K_d(\lambda, \alpha)K_d(-\lambda, \beta)K_d(\mu, \gamma) \\ &\qquad\qquad \times K_d\big(-\mu, -(\alpha + \beta + \gamma)\big)\, \mathbf{d}\alpha\, \mathbf{d}\beta\, \mathbf{d}\gamma\Big]. \end{aligned}$$

Now the last integral in the above expression multiplied by $d^{-3}c$ may, by Lemmas A.2 and A.1, be shown to be $O(d^{-1}\ln(N)\ln(d)\ln(c))$ under (C).

With the same method we obtain

$$d^{-3}cP_{d,c,T}(f, \lambda, \mu) = O\big(\ln(N)\ln(d)\ln(c)\big)d^{-2}\big(L^d\big)^2(\mu - \lambda).$$

Since for $\mu \not\equiv \pm\lambda \bmod(2\pi)$ and for d sufficiently large we have $(L^d)^2(\mu - \lambda) = |\mu - \lambda|^2$, it remains to prove that

$$d^{-3}c(2\pi)^2\widetilde{f}_d^2(\lambda)P_{d,c,T}(f, \lambda, \lambda) = \theta\big(dc^{-1}\big) + O(R_1).$$

To prove this, let $g_d(\lambda)$ be a sequence of AR(p_d) spectral densities, $p_d := d^{1/(1+r+\alpha)}$, r, α as in (B) with $\|g_d - f\|_\infty = O(p_d^{-(r+\alpha)})$. Then arguing as above and using Lemma A.3(b), we obtain

$$d^{-3}c|P_{d,c,T}(f, \lambda, \lambda) - P_{c,d,T}(g_d, \lambda, \lambda)| = O\big(d^{-\beta}\ln(N)\ln(d)\ln(c)\big),$$
$$\beta := \frac{r + \alpha}{1 + r + \alpha}.$$

Now denote $K^*_{d,p}(f,\lambda,\mu) := (2\pi)^{-1}\Sigma_{j=p}^{d-1}\overline{\phi}_j(\lambda)\phi_j(\mu)$, $\{\phi_k(\lambda)\}_{k\in N}$ as in Section 2 and define $P^*_{d,p,c,T}(f,\lambda,\mu)$ as $P_{d,c,T}(f,\lambda,\mu)$ by substituting $K^*_{d,p}(\cdot,\cdot)$ instead of $K_d(\cdot,\cdot)$. Again, (with $p := p_d$) we obtain

$$d^{-3}c|P^*_{d,p_d,c,T}(g_d,\lambda,\lambda) - P_{d,c,T}(g_d,\lambda,\lambda)| = O\big(d^{-\beta}\ln(N)\ln(d)\ln(c)\big).$$

Thus it remains to prove that

$$d^{-3}c(2\pi)^2 g_d^2(\lambda)P^*_{d,p_d,c,T}(g_d,\lambda,\lambda) = \theta\big(dc^{-1}\big) + O(R_1).$$

But in $P^*_{d,p_d,c,T}(g_d,\lambda,\lambda)$ enter only the polynomials orthogonal to g_d of degree $k \geq p_d$, which are exactly known [care of (A.2.4)]. By substituting them, one gets the desired result. □

Finally, we prove the lemma concerning the higher-order cumulants of $I^*_{d,c,T}(\lambda)$.

PROOF OF LEMMA 6. The proof follows the lines of the proof of Lemma 4.5 of Dahlhaus (1985). Let $\mathcal{P}^{(r)} = \{P_1,\dots,P_S\}$ be a fixed partition of the $2\times r$ table and let $s_i := |P_i|$. According to Corollary 1 and Lemma A.2 it is sufficient to show

$$\int\prod_{j=1}^{r}\big[L^d(\lambda_j+\alpha_j)L^d(-\lambda_j+\beta_j)L^N(c\alpha_j+c\beta_j)\big]\prod_{i=1}^{S}\mathbf{d}\widetilde{\kappa}_i$$
$$= O\Big(Nd^S\big(d/c\ln(N)\ln(c)\big)^{r-1}\ln(d)^{r-S+1}\Big),$$

since $S \leq r$.

To prove this, we assume without loss of generality that the P_j, $j = 1,\dots,S$, are enumerated in such a way that for each P_j there exists a P_k (for some $k<j$) and a row of the table, such that P_j and P_k contain at least one element of this row. This is possible by the indecomposability of the partition. Now denote

$$U_t := \{\kappa\in P_q, q>t,\ \text{such that } \kappa \text{ has a row neighbor } \in P_{q'}, q'\leq t\}$$

and

$$V_t := \{\kappa\in P_t \text{ such that } \kappa \text{ has a row neighbor } \in P_t\}$$

(where κ stands for some α or β of the $2\times r$ table) and let

$$n_t := \frac{|V_t|}{2},\qquad m_t := |U_{t-1}\cap P_t|,\qquad m_1 := 0.$$

We claim that:

(3.6) for $t \leq S$ the integral with respect to $\widetilde{\kappa}_i$, $i = 1,\dots,t$, of terms in the expression to be bounded, involving these variables is less than or equal to

$$O(1)L^N\Bigg(-\sum_{k\in U_t}c\kappa\Bigg)\prod_{i:=1}^{t}d\big[d/c\ln(N)\ln(c)\big]^{(s_i-1)-(m_i-1)^+-n_i}\big[\ln(d)\big]^{(m_i-1)^++n_i}.$$

To see that the desired result follows from (3.6), observe that:

(a) $(m_1 - 1)^+ = m_1 = 0, m_t \geq 1$ for $t \geq 2$, since $U_{t-1} \cap P_t \neq \emptyset$. These imply $\Sigma_{i=1}^{t}(m_i - 1)^+ = \Sigma_{i=2}^{t}(m_i - 1)$.

(b) $\Sigma_{t=1}^{S} n_t + \Sigma_{t=2}^{S} m_t = r$, since the first sum equals the number of rows occupied by elements of the same partition subset and the second, that of rows occupied by elements of different partition subsets.

Thus

$$\sum_{t=1}^{S} n_t + \sum_{t=1}^{S}(m_t - 1)^+ = r - S + 1.$$

These in turn yield a final bound of

$$\begin{aligned} &Nd^S\big[d/c\ln(N)\ln(c)\big]^{\Sigma_{i=1}^{S}[(s_i-1)-(m_i-1)^+-n_i]}\big[\ln(d)\big]^{\Sigma_{i=1}^{S}[(m_i-1)^+ + n_i]} \\ &\quad \leq Nd^S\big[d/c\ln(N)\ln(c)\big]^{r-1}\ln(d)^{r-S+1}. \end{aligned}$$

To check the validity of (3.6), it is sufficient to prove it: (a) for $t = 1$ and (b) for $t + 1$, assuming it holds for t. Set $P_t := \{\kappa_1, \dots, \kappa_{s_t}\}$. We indicate the proof of (b) assuming further that $\kappa_1 \in U_{t-1}, \kappa_{s_t} \notin U_{t-1}$ (note that the chosen enumeration of the partition subsets guarantees $U_{t-1} \cap P_t \neq \emptyset$). The other case (which may be treated by similar arguments) is that $P_t \subseteq U_{t-1}$. Assuming $\kappa_1 = a_p, \kappa_{s_t} = b_q$, Lemma A.1(ii) yields

$$\begin{aligned} &\int L^d(\lambda_d + \alpha_p)L^d\left(-\lambda_q - \alpha_p - \sum_{j=2}^{s_t-1}\kappa_j\right)L^N\left(-\sum_{\kappa\in U_{t-1}} c\kappa\right) \\ &\quad \times L^N\left(c\alpha_q - c\alpha_p - \sum_{j=2}^{s_t-1}c\kappa_j\right)\mathbf{d}\alpha_p \leq O\big(d/c\ln(N)\ln(c)\big) \\ &\quad \times L^d\left(\lambda_p - \lambda_q - \sum_{j=2}^{s_t-1}\kappa_j\right)L^N\left(c\alpha_q - \sum_{\kappa\in U_{t-1}\setminus P_t} c\kappa + \sum_{\kappa\in P_t\setminus U_{t-1}} c\kappa\right). \end{aligned}$$

Note that the $\kappa \in U_{t-1} \cap P_t$ do not appear any more in the argument of an L^N factor [we call this (*)]. Therefore one can integrate with respect to them by using Lemma A.1(i). After having integrated with respect to a $\kappa \in V_t$, (*) holds also for its row neighbor. So the number of times (*) occurs is $(m_t - 1)^+ + n_t$. The proof is completed by successively integrating with respect to all $\kappa \in P_t$, bearing in mind the above remark. It follows that the power of the factor $\ln(d)$ in the bound will be $(m_t - 1)^+ + n_t$, and the power of the factor $d/c\ln(N)\ln(c)$ will be $s_t - 1 - (m_t - 1)^+ - n_t$. □

3.5. *Proof of Theorem* 1. From Lemmas 2 and 3 it follows that $\sqrt{T/d}(\widehat{f}_{d,c,T} \times \widetilde{f}_d^{-1})(\lambda)$ and $\sqrt{T/d}I^*_{d,c,T}(\lambda)$ are asymptotically equivalent. On the other hand,

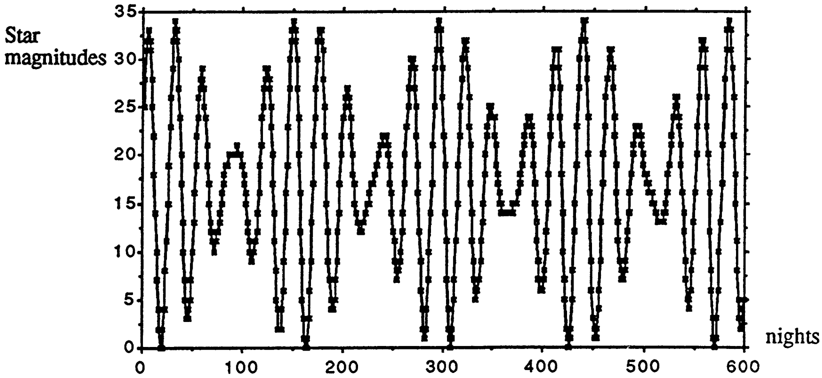

FIG. 1. *The data.*

using the cumulant method together with Lemmas 5 and 6, we ascertain that $\sqrt{T/d}[I^*_{d,c,T}(\lambda) - 1]$ is asymptotically normal and has the desired covariance structure. These prove part (a). From part (a) it follows that

$$\sqrt{Td^{-1}}\left\{\frac{\widehat{f}_{d,c,T}}{f}(\lambda_k) - 1 - B_d(\lambda_k)\right\}_{k=1,\dots,K} \Rightarrow_D \zeta^v \quad \text{where } B_d := \widetilde{f}_d f^{-1} - 1.$$

To see this, observe that using Lemma 1 we obtain

$$\begin{aligned}\widehat{f}_{d,T} f^{-1} - 1 - B_d &= \widetilde{f}_d f^{-1}\left[\widehat{f}_{d,T}\widetilde{f}_d^{-1} - 1\right] \\ &= \left[\widehat{f}_{d,T}\widetilde{f}_d^{-1} - 1\right]\left(1 + O\left(d^{-[(r+\alpha)\wedge 1]}\ln^2(d)\right)\right).\end{aligned}$$

Finally, part (b) is a direct consequence of the above and Lemma 1. □

4. Whittaker's variable-star example. In this section an example of "real" data is used to demonstrate the performance of the Capon estimator when compared to some other spectral estimators.

The data set we use represents the magnitudes of a variable star at midnight on 600 successive nights [Whittaker and Robinson (1924), page 349] (see Figure 1). It is the data set used by Bloomfield (1976) to demonstrate the leakage effect and the performance of the tapered periodogram [Bloomfield (1976), Sections 5.1–5.3]. Bloomfield suggests that "the data set consists approximately of the sum of two sinusoidal components" with frequencies $2\pi/29$ and $2\pi/24$, corresponding respectively to periods of 29 and 24 days. He also discusses extensively further properties of the data set, for example, the appearance of the higher harmonics of the two main peaks in the spectrum (see Figures 2 to 4); we therefore refer to him for this discussion.

All estimators are hereafter computed after subtracting the empirical mean from the original data. They are computed at the points $\lambda_k := 2\pi/n, k = 0, \dots, n-1$. We chose $n = 1392$ in order that $2\pi/29$ and $2\pi/24$ belong to the grid. The

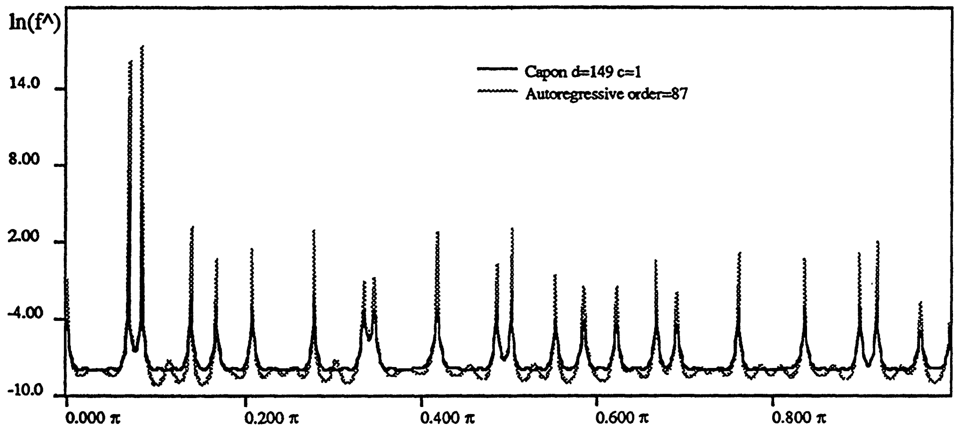

FIG. 2. *The Capon and the autoregressive estimator.*

estimators are plotted on a logarithmic scale (natural logarithm). [The computations were carried out on a Macintosh II and for all of them double precision was used (mantissa length = 64).]

The estimators shown in Figures 2 to 4 are the following:

(a) The Capon estimator as defined in (1.1) and (1.2). According to Remark 1 we have chosen the parameter $c = 1$. The parameter value $d = 149$ was chosen based on the criterion described in Remark 3.

(b) The autoregressive least-squares estimator $\widehat{\phi}_p$ of order p defined as follows: let $\mathbf{a}_p^t := (1, a_{p,1}, \dots, a_{p,p})$ and define the residuals $e_{p,t}(\mathbf{a}) := X_t + a_{p,1}X_{t-1} + \cdots + a_{p,p}X_{t-p}$, $t = p+1, \dots, T$. Further define $\widehat{\sigma}_p^2 := \inf_{\mathbf{a}} \Sigma_{t=p+1}^{T} e_{p,t}^2(\mathbf{a})$, the

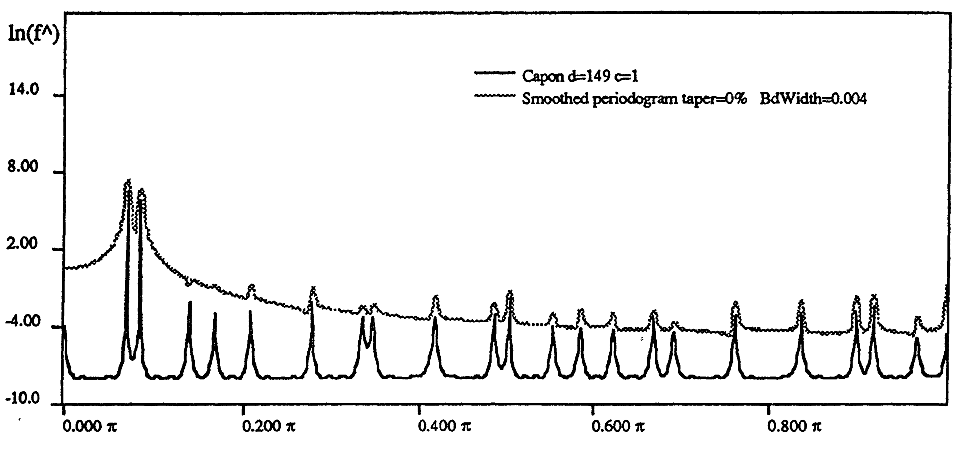

FIG. 3. *The Capon and the untapered smoothed periodogram.*

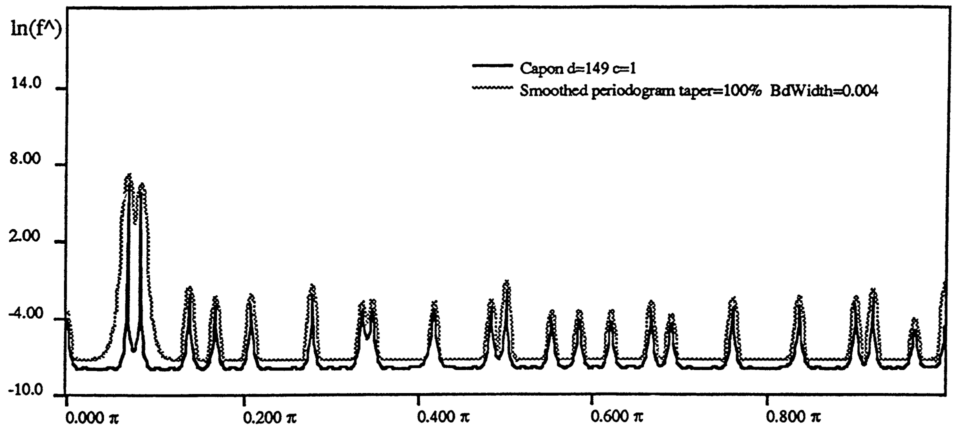

FIG. 4. *The Capon and the tapered smoothed periodogram.*

infimum being attained at $\widehat{\mathbf{a}}_p$. Then $\widehat{\phi}_p(\lambda) := (\widehat{\sigma}_p^2/2\pi)|\widehat{\mathbf{a}}_p^t b_\lambda|^{-2}$. The order $p = 87$ was chosen by the AIC.

(c) The periodogram $\widehat{f}_{p,b}$ $p\%$ tapered and smoothed with bandwidth b, defined as follows: let $u_{p,t}, t = 1,\dots,T$, be the $p\%$ Tukey data taper, defined, for example, in Bloomfield (1976), Section 5.2 (our p corresponds to $2m/n$ in Bloomfield's notation) and further, $I_p^T(\lambda) := (2\pi U_T)^{-1}|\Sigma_{t=1}^T u_{p,t} X_t e^{i\lambda t}|^2$, where $U_T := \Sigma_{t=1}^T u_{p,t}^2$. Then $\widehat{f}_{p,b}(\lambda_k) := w^{-1}\Sigma_{|j|\le Tb/2} w_j I_p^T(\lambda_{k-j})$, where w is the sum of the w_j's entering in the previous mean and $w_j := W(\lambda_j/b)$, W being the Bartlett–Priestley kernel: $W(x) := 1-(x/\pi)^2$ on $(-\pi,\pi)$ and 0 elsewhere. In Figures 3 and 4 we show the periodograms 0% and 100% tapered, smoothed with bandwidth $b = 0.004$, which was chosen empirically.

In order to indicate the uncertainty of the four estimators, we calculate the standard deviations of their (natural) logarithms (for the parameter values chosen):

(a) For the Capon estimator it equals approximately (Theorem 1): $[2/3\, d/(T - d + 1)]^{1/2} = 0.47$.

(b) For the autoregressive estimator it equals approximately: $[p/(T - p + 1)]^{1/2} = 0.41$.

(c) For the nontapered, smoothed periodogram it equals approximately $[w^{-2}\Sigma_{|j|\le Tb/2} w_j^2]^{1/2} = 0.68$, and for the tapered one it equals approximately the same quantity multiplied by $(T\Sigma_{t=1}^T u_{p,t}^4)^{1/2}/\Sigma_{t=1}^T u_{p,t}^2$ which gives in total 0.94.

The differences of the four estimators in Figures 2 to 4 are much larger than one would expect from the calculated standard deviations. This is due to the differences in their bias.

DISCUSSION. From Figure 2 it is clear that the untapered periodogram suffers from leakage and fails to discover clearly the higher harmonics of the two main peaks. The other three estimators have in general the same shape. Observe that the Capon estimator and the autoregressive LS estimator do not suffer at all from leakage, although no taper was used for them. Moreover, one has the impression that they distinguish the two main peaks more clearly than the periodograms. Besides there is a difference between the autoregressive LS estimator on the one side and the Capon estimator and the periodograms on the other, that is, in the estimated ordinates of the spectrum in the peaks. Our suggestion is that the difference is due rather to the larger bias of the Capon estimator and the periodograms than to random fluctuation of the autoreggressive LS estimator (see Remark 4).

APPENDIX

A.1. The L^N functions. We introduce the $L^N(\lambda)$ functions [Dahlhaus (1983)] and state some of their properties which were used for the cumulant calculations.

Let $L^N(\lambda)$ be a 2π-periodic and symmetric around 0 function, defined on $[-\pi,\pi]$ as follows:

$$L^N(\lambda) = \begin{cases} N, & \text{if } |\lambda| \le N^{-1}, \\ |\lambda|^{-1}, & \text{elsewhere.} \end{cases}$$

Then the following lemma holds.

LEMMA A.1. *There exists a $K \in \mathbf{R}$, such that for $N, d, c \in \mathbf{N}^+$, $\alpha, \beta, \gamma, \lambda, \mu, x \in \mathbf{R}$:*

(i) $\int L^N(\gamma+\alpha)\, L^N(\beta-\alpha)\, \mathbf{d}\alpha \le K \ln(N) L^N(\gamma+\beta)$ [*Dahlhaus* (1983)].

(ii) *If $c/d \le C$ for some C holds, then there exists $K' \in \mathbf{R}$, such that*

$$\int L^d(\gamma+x)L^d(\beta-x)L^N(c\lambda+cx)L^N(c\mu-cx)\,\mathbf{d}x$$
$$\le K' d c^{-1} \ln(N)\ln(c) L^d(\gamma+\beta) L^N(c\lambda+c\mu) \quad [\textit{Dahlhaus } (1985)].$$

A.2. Orthogonal polynomials. In this section we study some properties of quantities related to the system $\{\phi_k(\lambda)\}_{k\in N}$ of polynomials orthogonal with respect to a spectral density f, where f fulfills assumptions (A) and (B) [see also Szegö (1959)]. We study especially bounds and approximations of kernels of the type $K_d(\lambda,\mu) = K_d(f,\lambda,\mu)$ defined in Section 3. Let also Γ_d and U_d be defined as in Section 1.

Let $\{\phi_k(\lambda)\}_{k\in N}$ denote the system of polynomials orthogonal with respect to f, that is, ϕ_k is a polynomial of degree k in $e^{i\lambda}$ and $(2\pi)^{-1}\int f(\lambda)\phi_i(\lambda)\overline{\phi_j(\lambda)} = \delta_{ij}$.

Then it may be seen that

$$\phi_d(\lambda) = \sqrt{2\pi} e_d^t U_d^{-1} b_\lambda, \qquad e_d^t := (0, \dots, 0, 1) \in \mathbf{R}^d, \tag{A.2.1}$$

$$K_d(\lambda, \mu) = (2\pi)^{-1} \sum_{i=0}^{d-1} \overline{\phi_i(\lambda)} \phi_i(\mu). \tag{A.2.2}$$

From Szegö (1959), Theorem 12.1.3, it is further well known that under assumption (A) (in Section 2):

$$|\phi_d(\lambda)|^2 \to f^{-1}(\lambda) \quad \text{uniformly in } \lambda \text{ for } d \to \infty. \tag{A.2.3}$$

Moreover, if $f = |h_p|^{-2}$, where h_p is a polynomial of degree p in $e^{i\lambda}$, then for $j \geq p$,

$$\phi_j(\lambda) = \overline{h_p(\lambda) e^{i\lambda j}}. \tag{A.2.4}$$

Finally, the Christoffel–Darboux [Szegö (1959), Theorem 11.4.2] formula holds

$$K_d(\lambda, \mu) = (2\pi)^{-1} \frac{e^{id(\mu - \lambda)} \phi_d(\lambda) \overline{\phi_d(\mu)} - \overline{\phi_d(\lambda)} \phi_d(\mu)}{1 - e^{i(\mu - \lambda)}}. \tag{A.2.5}$$

At first one may bound the $K_d(f, \lambda, \mu)$ kernels by an L_d function as follows.

LEMMA A.2. *Let f fulfill* (A). *Then there exists a constant M', which depends only on m, M [from* (A)], *such that*

$$K_d(f, \lambda, \mu) \leq M' L^d(\lambda - \mu).$$

PROOF. Cauchy's inequality yields: $|\overline{b}_\lambda^t \Gamma_d^{-1} b_\mu|^2 \leq \overline{b}_\lambda^t \Gamma_d^{-1} b_\lambda \overline{b}_\mu^t \Gamma_d^{-1} b_\mu \leq d^2 m^{-2}$. On the other hand, from (A.2.5) we have

$$\left|\overline{b}_\lambda^t \Gamma_d^{-1} b_\mu\right|^2 \leq 2(2\pi)^{-1} |\phi_d(\lambda)|^2 |\phi_d(\mu)|^2 \left(1 - \cos(\mu - \lambda)\right)^{-1}.$$

Because of (A.2.3), the result follows from $x \sin^{-1}(x/2) \leq M'$, $x \in [-\pi, \pi]$. □

In the following lemma the orthogonal polynomials corresponding to f are approximated by the orthogonal polynomials corresponding to an AR(p) spectral density approximating f [see also Hannan and Wahlberg (1989)].

LEMMA A.3. *Let f fulfill conditions* (A) *and* (B). *Then for a given sequence of integers $p_n \to \infty$ with $p_n \leq n$ there exists a sequence $|t_n|^2$ of positive trigonometric polynomials of degree p_n, such that*

(a) $\| |t_n|^2 - f^{-1}\|_\infty = O(p_n^{-\gamma})$ *and* $\|\phi_n - \psi_{n,n}\|_\infty = O(p_n^{-\gamma}\ln(n))$, *where* $\{\psi_{n,k}(\lambda)\}_{k\in N}$ *denotes the system of orthogonal polynomials associated with* $|t_n|^{-2}$ *and* $\gamma := r + \alpha$, r, α *as in* (B).

(b) $\sup_{\lambda,\mu\in R} |K_n(f, \lambda, \mu) - K_n(|t_n|^{-2}, \lambda, \mu)| = O(p_n^{-\gamma}\ln(n))L^n(\mu - \lambda)$.

PROOF. From Butzer and Nessel (1971), Theorem 2.2.3, it follows that (for n big enough) there exists a sequence $|t_n|^2$ of positive trigonometric polynomials of degree p_n with $\| |t_n|^2 - f^{-1}\|_\infty = O(p_n^{-\gamma})$. We will show that this sequence also fulfills (a) and (b).

Denote the $n \times n$ Toeplitz matrix associated with $|t_n|^{-2}$ by G_n and let $G_n = V_n V_n^t$ be its Cholesky decomposition. First observe that $\|\Gamma_n - G_n\| \le 2\pi \| |t_n|^{-2} - f\|_\infty$ and further that $\|\Gamma_n^{-1} - G_n^{-1}\| = O(p_n^{-\gamma})$. To see the latter, observe that $\Gamma_n^{-1} - G_n^{-1} = (U_n^t)^{-1}[I - (U_n^{-1}G_n(U_n^t)^{-1})^{-1}]U_n^{-1}$ and further that a Neumann expansion yields

$$\left\|I - \left(U_n^{-1}G_n\left(U_n^t\right)^{-1}\right)^{-1}\right\| = \left\|I - \left(U_n^{-1}G_n\left(U_n^t\right)^{-1}\right)\right\|$$
$$= O\left(\|\Gamma_n - G_n\|\right) = O\left(p_n^{-\gamma}\right).$$

To prove (a), let us expand ϕ_n with respect to $\{\psi_{n,k}(\lambda)\}_{k\in N}$. We obtain

$$\phi_n(\lambda) = \alpha_n \psi_{n,n}(\lambda) + \int \left[|t_n|^{-2} - f\right](\mu) \sum_{\nu=0}^{n-1} \psi_{n,\nu}(\lambda)\overline{\psi_{n,\nu}}(\mu)\phi_n(\mu)\,\mathbf{d}\mu$$
$$+ \int f(\mu) \sum_{\nu=0}^{n-1} \psi_{n,\nu}(\lambda)\overline{\psi_{n,\nu}}(\mu)\phi_n(\mu)\,\mathbf{d}\mu,$$

with $\alpha_n = \int |t_n|^{-2}\overline{\psi_{n,n}}(\mu)\phi_n(\mu)\,\mathbf{d}\mu$. The last term vanishes because of the orthogonallity property of ϕ_n with respect to f. The second term on the right side is less than or equal to [note (A.2.3) and Lemma A.2]

$$\left\| |t_n|^{-2} - f\right\|_\infty \sup_{\lambda,n} |\phi_n(\lambda)| \int L^n(\lambda)\,\mathbf{d}\lambda = O\left(p_n^{-\gamma}\ln(n)\right).$$

Thus for (a) it is sufficient to show $|\alpha_n - 1| = O(p_n^{-\gamma})$. Now from (A.2.1):

$$\alpha_n = \int |t_n|^{-2}\overline{\psi_{n,n}}(\mu)\phi_n(\mu)\,\mathbf{d}\mu = e_n^t U_n^{-1}G_n(V_n^t)^{-1}e_n = e_n^t U_n^{-1}V_n e_n$$
$$= \left(U_n^{-1}\right)_{nn}(V_n)_{nn},$$

where we denote the n, nth element of the matrix A by $(A)_{nn}$.

Since $(V_n^{-1})_{nn} = (V_n)_{nn}^{-1}$ and $(V_n^{-1})_{nn}^2 = (G_n^{-1})_{nn} \ge (\pi m)^{-1}$ it is sufficient to show

$$\left(U_n^{-1}\right)_{nn} - \left(V_n^{-1}\right)_{nn} = O\left(p_n^{-\gamma}\right).$$

But this follows from

$$(U_n^{-1})^2_{nn} = (\Gamma_n^{-1})_{nn}, \qquad (V_n^{-1})^2_{nn} = (G_n^{-1})_{nn},$$
$$\|\Gamma_n^{-1} - G_n^{-1}\| = O(p_n^{-\gamma}) \quad \text{and} \quad \|\Gamma_n^{-1}\| \geq 2\pi m^{-1}.$$

This completes the proof of (a).

Let us now prove (b). First we obtain from Cauchy's inequality

$$|\bar{b}_\lambda^t \Gamma_n^{-1} b_\mu - \bar{b}_\lambda^t G_n^{-1} b_\mu| \leq n \|\Gamma_n^{-1} - G_n^{-1}\| = nO(p_n^{-\gamma}).$$

On the other hand, from (A.2.5) we obtain

$$|K_n(f, \lambda, \mu) - K_n(|t_n|^{-2}, \lambda, \mu)| \leq O(1)\|\phi_n - \psi_{n,n}\|_\infty |1 - \cos(\mu - \lambda)|^{-1/2}$$

and the result follows from (a) and from $x \sin^{-1}(x/2) \leq M'$, $x \in [-\pi, \pi]$. □

A.3. The variance function θ. Let θ be defined as in Section 2. The following lemma gives its relevant properties.

LEMMA A.4. *θ fulfills the following*:

(i) $\theta(x) \to \frac{2}{3}$, $x \to \infty$.
(ii) $\theta(x) = x^{-1}$, $0 < x \leq 1$.
(iii) $\theta(x) \geq \frac{2}{3}$, $x \geq 0$.
(iv) $\operatorname{arginf}\{\theta(dc^{-1}),\ c \in \mathbf{N}^+\} = 1$, $\forall d \in \mathbf{N}^+$.

PROOF. For $x \in \mathbf{R}^+$ let $\alpha_x := x - [x] \in [0, 1)$. Then θ may be written as

$$\theta(x) = \frac{2}{3} + \frac{x^{-2}}{3} - \left(\frac{2}{3}\alpha_x^3 - \alpha_x^2 + \frac{\alpha_x}{3}\right)x^{-3}. \tag{A.3.1}$$

If $x \in \mathbf{N}$, then $\alpha_x = 0$. We obtain

$$\theta(x) = \frac{2}{3} + \frac{x^{-2}}{3}. \tag{A.3.2}$$

If $x \leq 1$, then $\alpha_x = x$ and $\theta(x) = x^{-1}$. This proves (ii).

(A.3.3) For $k \leq x < k + 1$ for some k it is easily seen that θ is decreasing.

Let us now prove (iv). We have for $c \in N^+$ and $x := dc^{-1}$:

$$3d^3\left[\theta(d) - \theta(dc^{-1})\right] = c^3(2\alpha_x^3 - 3\alpha_x^2 + \alpha_x) - c^2 d + d \leq Ac^3 - c^2 d + d,$$

with

$$A := \sup_{0 \leq \alpha \leq 1}(2\alpha^3 - 3\alpha^2 + \alpha) \leq \sqrt{3}/12.$$

Now for fixed d and $c \leq d$ the function $Ac^3 - c^2 d + d$ is falling in c and is negative for $c = 2$, thus it is negative for $2 \leq c \leq d$. This proves (iv). Also together with (A.3.2) it proves (iii) and further taking (A.3.3) into account, (i) follows as well. □

A.4. The nonzero mean case. In this section we indicate the proof of Theorem 1 in the nonzero mean case (see Remark 5). It is sufficient to show that

$$\sqrt{\frac{T}{d}}\|\widehat{\Gamma}_{d,T} - \widehat{\Gamma}^{\circ}_{d,T}\| \to_P 0, \tag{A.4.1}$$

where $\widehat{\Gamma}_{d,T}$ and $\widehat{\Gamma}^{\circ}_{d,T}$ are defined as in (1.2) but are based on $X_i - \mu$ and $X_i - \overline{X}_T$, respectively. To see how Theorem 1 follows from (A.4.1), observe that from (A.4.1) we obtain for the estimator based on the empirically centered data the same expansion as in Lemma 3; (A.4.1) further ascertains that the difference of the first-order expansion terms $I^*_{d,c,T}$ for the two estimators (based on $X_i - \mu$ and $X_i - \overline{X}_T$, respectively) tends stochastically to 0 when blown up by $\sqrt{T/d}$.

To prove (A.4.1), assume for simplicity that $c = 1$. Then the difference of the (t, s) element of the two matrices may be bounded by

$$\left|\widehat{\Gamma}_{d,T} - \widehat{\Gamma}^{\circ}_{d,T}\right|_{t,s} \le (\overline{X}_T - \mu)^2(1 + dT^{-1}) + dT^{-1}|\overline{X}_T - \mu| L_{d,T}, \tag{A.4.2}$$

where

$$L_{d,T} := \sup_{n=1,\dots,d} d^{-1}\left[\sum_{i=1}^{n-1}(X_i - \mu) + \sum_{i=T-d+n+1}^{T}(X_i - \mu)\right].$$

We now state further assertions, which will be proven below. Under the assumptions of Theorem 1, we have

$$|\overline{X}_T - \mu| = O_P(\ln(T)^{1/2}T^{-1/2}), \tag{A.4.3}$$

$$\text{for any } \varepsilon > 0: \quad L_{d,T} = O_P(d^{-(1/2-\varepsilon)}). \tag{A.4.4}$$

Assertion (A.4.1) follows from (A.4.2), (A.4.3) and (A.4.4): setting E_d to be the $d \times d$ matrix filled with "1," we obtain

$$\begin{aligned}\sqrt{\frac{T}{d}}\|\widehat{\Gamma}_{d,T} - \widehat{\Gamma}^{\circ}_{d,T}\| &\le \sqrt{\frac{T}{d}}\|E_d\|\left[O_P\left(\frac{\ln(T)}{T}\right) + O_P\left(\frac{d\ln(T)^{1/2}L_{d,T}}{T^{3/2}}\right)\right] \\ &= O_P\left(\ln(T)\sqrt{\frac{d}{T}} + \ln(T)^{1/2}\frac{d^{1+\varepsilon}}{T}\right) \to_P 0.\end{aligned}$$

It remains to prove (A.4.3) and (A.4.4). They will both follow from (A.4.5) below: for any $k \in Z^+$ we have

$$E\left[\sum_{i=1}^{T}(X_i - \mu)\right]^{2k} = O(T^k \ln(T)^k). \tag{A.4.5}$$

We first prove (A.4.5). Assuming without loss of generality that $\mu = 0$, we obtain from Proposition 1(a) and Lemma A.2 (using the notation of the $2 \times k$ table and with E_T as above):

$$E\left[\sum_{i=1}^{T} X_i\right]^{2k} = E[\boldsymbol{X}^T E_T \boldsymbol{X}]^k = O(1)\sum_{\mathrm{ap},(k)}\int\prod_{i=1}^{k} L^T(\alpha_i)L^T(\beta_i)\prod_{i=1}^{S} d\widetilde{\kappa}_i.$$

Applying Lemma A.1(i) successively, one gets that this quantity is $O(T^S \times \ln(T)^{2k-S})$. This is sufficient for (A.4.5), since $S \leq K$ (because partitions containing one-element subsets have contribution equal to 0).

The proof of (A.4.3) follows by applying Chebyshev's inequality and using (A.4.5), with k chosen appropriately.

$$P\left[d^{1/2-\varepsilon}L_{d,T} \geq M\right] \leq 2dd^{-2k}\left(\frac{d^{1/2-\varepsilon}}{M}\right)^{2k}\left[d\ln(d)\right]^k = O(1)\ \ln(d)^k d^{1-2k\varepsilon},$$

which tends to 0, if we chose $k > 1/(2\varepsilon)$. The proof of (A.4.4) follows similarly. □

Acknowledgments. Here we would like to thank Professor R. Dahlhaus, Professors P. Doukhan and P. D. W. Ehm, as well as the referees whose comments substantially contributed to the contents and the appearance of this paper.

REFERENCES

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

BLOOMFIELD, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

BRILLINGER, R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

BURG, J. P. (1972). The relationship between maximum entropy and maximum likelihood spectra. *Geophysics* **37** 375–376.

BUTZER, P. and NESSEL, R. (1971). *Fourier Analysis and Approximation,* **1***: One Dimensional Theory.* Birkhäuser, Boston.

BYRNE, C. L. and FITZGERALD, R. M. (1984). Spectral estimators that extend the maximum entropy and the maximum likelihood methods. *SIAM J. Appl. Math.* **44** 425–442.

CAPON, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57** 1408–1418.

CAPON, J. and GOODMAN, N. R. (1970). Probability distributions for estimates of the frequency-wavenumber spectrum. *Proc. IEEE* **58** 1785–1786.

CHAVE A. D., DOUGLAS S. L. and FILLOUX, J. H. (1991). Variability of the wind stress curl over the North Pacific: implications for the oceanic response. *Journal of Geophysical Research* **96** 18361–18379.

DAHLHAUS, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4** 163–175.

DAHLHAUS, R. (1985). On a spectral density estimate obtained by averaging periodograms. *J. Appl. Probab.* **22** 598–610.

HANNAN, E. J. and WAHLBERG, B. (1989). Convergence rates for inverse Toeplitz matrix forms. *J. Multivariate Anal.* **31** 127–135.

HURVICH, C. M. and TSAI, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76** 297–307.

LEWIS, R. and REINSEL, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* **16** 393–411.

MARZETTA, T. L. (1983). A new interpretation for Capon's maximum likelihood method of frequency-wavenumber spectral estimation. *IEEE Trans. Acoust. Speech Signal Process.* **31** 445–449.

MCDONOUGH, R. N. (1979). Application of the maximum likelihood method and the maximum entropy method to array processing. In *Nonlinear Methods of Spectral Analysis* (S. Haykin, ed.) 181–245. Springer, Berlin.

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

PISARENKO, V. F. (1972). On the estimation of spectra by means of non-linear functions of the covariance matrix. *Geophysical Journal of the Royal Astronomical Society* **28** 511–531.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston.

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.

SUBBA RAO, T. and GABR, M. M. (1989). The estimation of spectrum, inverse spectrum and inverse autocovariances of a stationary time series. *J. Time Ser. Anal.* **10** 183–202.

SZEGÖ, G. (1959). *Orthogonal Polynomials.* Amer. Math. Soc., New York.

WHITTAKER, E. T. and ROBINSON, G. (1924). *The Calculus of Observations*. Blackie and Son, London.

YOSIDA, K. (1980). *Functional Analysis*, 6th ed. Springer, Berlin.

ZHURBENKO, I. G. (1980). On the efficiency of spectral density estimates of a stationary process. *Theory Probab. Appl.* **25** 466–480.

12, RUE DE L' ECHIQUIER
75010 PARIS
FRANCE