Metrika (2007) 65: 133–157
DOI 10.1007/s00184-006-0065-8

ORIGINAL ARTICLE

Michael Eichler

A frequency-domain based test for non-correlation between stationary time series

Received: 30 March 2005 / Published online: 2 June 2006

Abstract A one-sided asymptotically normal test for non-correlation between two stationary time series is proposed based on the spectral coherence function. The test statistic is a properly standardized version of the integrated spectral coherency and has similar asymptotic properties as a previously introduced time domain based test for non-correlation. Unlike its time domain counterpart, the proposed test does not require prewhitening of the time series and, thus, is a truly nonparametric test for non-correlation. In a simulation study, we evaluate the small sample performance of the proposed test in comparison with the time domain test and address the problem of bandwidth selection. Furthermore, we present a modification of the test statistic that allows to test for non-correlation over frequency bands. This version shows higher power of detecting interrelationships restricted to the frequency band of interest.

Keywords Multivariate time series · Non-correlation · Spectral coherence · Nonparametric test · Bandwidth selection

1 Introduction

In multivariate time series analysis, testing for independence or non-correlation between two time series is an important problem that has been recently addressed

This work has been carried out at the Institute of Applied Mathematics at the University of Heidelberg and partly while the author was visiting the Department of Statistics at the University of Chicago.

M. Eichler
Department of Quantitative Economics,
University of Maastricht,
P.O.Box 616, 6200 MD Maastricht,
The Netherlands
E-mail: m.eichler@ke.unimaas.nl

in many papers. Most tests that have been developed are for finite order vector autoregressive (VAR) or vector autoregressive moving average (VARMA) processes (e.g., Haugh 1976; Koch and Yang 1986; El Himdi and Saidi 1997; Hallin and Saidi 2005). More generally, Hong (1996) considers infinite order autoregressive series and proposes first fitting autoregressive models to each series and then taking a weighted sum of the squared residual cross-correlations. In contrast to Haugh's test, the order of the fitted autoregression increases with the sample size and, thus, protects against misspecification of the true VARMA model at least for large enough sample sizes. A robust version of Hong's test for ARMA series is presented in Duchesne and Roy (2003).

Alternatively, one might take a frequency domain approach and test for non-correlation using the spectral coherence, which can be estimated nonparametrically using kernel estimates for the (cross) spectral densities. The frequency domain approach to time series analysis has been discussed in great detail in Brillinger (1981). Tests for non-correlation in the frequency domain so far have been based on pointwise test bounds. Dahlhaus et al. (1997) proposed to use the maximum of the spectral coherence and provided heuristic approximate test bounds.

In this paper, we consider the integrated spectral coherence, which is the frequency domain version of Hong's test. Unlike the latter, the frequency domain approach does not require prewhitening of the two time series and thus provides a completely nonparametric approach for testing for non-correlation.

In section 2, we introduce the test statistic. The asymptotic distribution under the null hypothesis and under a series of local hypothesis is discussed in section 3, while in section 4 we investigate the small sample properties of the proposed test statistic by simulation methods. In particular, we show that the frequency domain approach yields better results than the time domain version if the time series exhibit strong periodicities. The results also emphasize the need for selecting an appropriate bandwidth for smoothing the periodogram. The problem of data-driven bandwidth selection is briefly discussed in section 5. Finally, in section 6, we present a simple modification to test for non-correlation over a fixed frequency band. The methods are illustrated by an application to neurological data. The mathematical proofs of the results are technical and put into the appendix.

2 The test statistic

Let $(X_t) = (X_{t,1}, X_{t,2})$ be a bivariate stationary process with mean zero and covariances $c_{ab}(u) = \mathbb{E}(X_{t,a} X_{t-u,b})$ such that the (cross) spectral densities

$$f_{ab}(\lambda) = \frac{1}{2\pi} \sum_{u \in \mathbb{Z}} c_{ab}(u) \exp(-\mathrm{i}\lambda u), \quad \lambda \in [-\pi, \pi]$$

exist for $a, b = 1, 2$. Then it is well known that $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated if and only if the cross spectral density $f_{12}(\lambda)$ is zero for all frequencies $\lambda \in [-\pi, \pi]$ or, equivalently, if and only if the spectral coherence

$$|R_{12}(\lambda)|^2 = \frac{|f_{12}(\lambda)|^2}{f_{11}(\lambda)\, f_{22}(\lambda)} \tag{1}$$

vanishes for all $\lambda \in [-\pi, \pi]$. In other words, $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated if and only if all their frequency components are uncorrelated.

Given data $X_1, \ldots, X_T$, nonparametric estimation of the spectral coherence is based on the tapered periodogram

$$I_{ab}^{(T)}(\lambda) = \left(2\pi H_{2,T}\right)^{-1} d_a^{(T)}(\lambda)\, d_b^{(T)}(-\lambda),$$

where

$$d_a^{(T)}(\lambda) = \sum_{t=1}^{T} h_{t,T}\, X_{t,a}\, \exp(-\mathrm{i}\lambda t)$$

is the discrete Fourier transform of $(X_{t,a})$, $h_{t,T}$ is a data taper with $h_{t,T} = h(t/T)$ for some taper function h (e.g., Dahlhaus 1983), and $H_{2,T} = \sum_{t=1}^{T} h_{t,T}^2$. Then the spectral density $f_{ab}(\lambda)$ can be estimated by the kernel estimator

$$\hat{f}_{ab}(\lambda) = \sum_{j=0}^{T-1} w^{(T)}\left(\lambda - \frac{2\pi j}{T}\right) I_{ab}^{(T)}\left(\frac{2\pi j}{T}\right),$$

where $w^{(T)}$ is a 2π-periodic weight function with $w^{(T)}(\lambda) = M_T w(M_T \lambda)$ for $\lambda \in [-\pi, \pi]$ and some kernel function w (e.g., Brillinger 1981). M_T determines the effective number of frequencies over which the periodogram is averaged and should increase with the sample size T; furthermore, $B_T = M_T^{-1}$ is called the bandwidth of the kernel. An appropriate rate of increase for M_T will be specified in Theorem 5. Substituting the spectral estimators $\hat{f}_{ab}$ into (1), we obtain

$$|\hat{R}_{12}^{(T)}(\lambda)|^2 = \frac{|\hat{f}_{12}(\lambda)|^2}{\hat{f}_{11}(\lambda)\, \hat{f}_{22}(\lambda)}$$

as an estimate for the spectral coherence. Since the spectral coherence is nonnegative, we can take the estimated integrated spectral coherence

$$S_T = \frac{1}{T} \sum_{j=0}^{T-1} \left| \hat{R}_{12}^{(T)}\left(\frac{2\pi j}{T}\right) \right|^2$$

to test for non-correlation between $(X_{t,1})$ and $(X_{t,2})$. An appropriately standardized version of S_T is

$$Q_T = \frac{T\, S_T - M_T\, \mu}{\sqrt{M_T}\, \sigma}$$

with

$$\mu = \frac{H_4}{H_2^2} \int_{-\infty}^{\infty} W(\alpha)^2 d\alpha \quad \text{and} \quad \sigma^2 = \frac{2H_4^2}{H_2^4} \int_{-\infty}^{\infty} W(\alpha)^4 d\alpha, \tag{2}$$

where $H_k = \int_0^1 h(t)^k\, dt$ and $W(\alpha) = \int_{-\infty}^{\infty} w(\lambda)\, \exp(\mathrm{i}\lambda\alpha)\, d\lambda$.

We note that the test statistic Q_T is of similar form as the time domain based statistic proposed by Hong (1996), who suggested to test for non-correlation between $(X_{t,1})$ and $(X_{t,2})$ by first prewhitening $X_{t,1}$ and $X_{t,2}$ and then testing for non-correlation between the residuals, say $\hat{\varepsilon}_{t,1}$ and $\hat{\varepsilon}_{t,2}$, based on the statistic

$$S_T^* = \sum_{u=1-T}^{T-1} W\left(\frac{u}{M_T}\right)^2 \hat{\rho}_{\varepsilon_1\varepsilon_2}(u)^2,$$

where $W(u)$ is a lag window and $\hat{\rho}_{\varepsilon_1\varepsilon_2}(u)$ is the residual cross-correlation function obtained from the residuals $\hat{\varepsilon}_{t,1}$ and $\hat{\varepsilon}_{t,2}$. This approach leads to the test statistic

$$Q_T^* = \frac{T\, S_T^* - M_T\, \mu^*}{\sqrt{M_T}\, \sigma^*}, \tag{3}$$

where the constants μ^* and σ^* are given by (2) with $H_2 = H_4 = 1$. In the following section, we show that the proposed frequency domain test statistic has similar properties as the time domain version proposed by Hong (1996). However, the frequency domain approach presented here is truly nonparametric since it does not require prewhitening of the series.

An alternative nonparametric frequency-domain based test for non-correlation has been proposed by Dahlhaus et al. (1997), who considered the maximal spectral coherence

$$S_T^{**} = \max_{0 \le j < T} \left| \hat{R}_{12}^{(T)}\left(\frac{2\pi j}{T}\right) \right|^2.$$

A comparison of the definitions of S_T and S_T^{**} suggests that S_T, measuring the mean deviation from the null hypothesis, leads to more powerful tests if the spectral coherence differs from zero over a large range of frequencies. In contrast, tests based on S_T^{**} should be superior if the null hypothesis is violated only at a single frequency. We note that the exact asymptotic distribution of the maximal spectral coherence S_T^{**} has not yet been established and the test proposed by Dahlhaus et al. (1997) relies on a heuristic approximation.

3 Asymptotic properties

For the discussion of the asymptotic properties of the test statistic Q_T, we require the following mixing condition of the process (X_t).

Assumption 1 $(X_t) = (X_{t,1}, X_{t,2})$ is a bivariate stationary process with mean zero such that

$$\sum_{u_1,\ldots,u_{k-1} \in \mathbb{Z}} \big(1 + |u_j|\big)\, |c_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1})| < \infty$$

for all $k \ge 2$ and $1 \le j \le k-1$, where $c_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1})$ is the kth order cumulant of $X_{u_1,a_1}, \ldots, X_{u_{k-1},a_{k-1}}, X_{0,a_k}$.

This condition, which is a special version of Assumption 2.6.2 of Brillinger (1981), is satisfied, for example, for all ARMA processes of finite order. The assumption implies that the spectral densities exist and have bounded and uniformly continuous derivatives. Furthermore, we require that the spectral matrix is bounded away from zero.

Assumption 2 The spectral matrix $f(\lambda) = (f_{ab}(\lambda))$ is nonsingular for all frequencies $\lambda \in [-\pi, \pi]$.

Additionally, we impose the following assumptions on the taper function h and the kernel function w.

Assumption 3 The taper function $h(x)$ is a bounded real function of bounded variation with $h(x) = 0$ for $x \notin [0, 1]$.

The assumption in particular includes the nontapered case, where $h(x) = 1$ for $x \in [0, 1]$ and 0 otherwise. However, we note that the small sample properties of spectral estimates can be improved considerably by choosing a smooth taper function with $h(0) = h(1) = 0$ (e.g., Dahlhaus 1990). The small sample properties of the test statistic will be investigated by simulation methods in the next section.

Assumption 4 The kernel function $w(\lambda)$ is bounded, symmetric, nonnegative, and Lipschitz continuous with $\int_{-\infty}^{\infty} w(\lambda)\, d\lambda = 1$ and $\int_{-\infty}^{\infty} \lambda^2\, w(\lambda)\, d\lambda < \infty$. Furthermore $w(\lambda)$ has continuous Fourier transform $W(\alpha)$ such that $\int_{-\infty}^{\infty} W(\alpha)^2\, d\alpha < \infty$ and $\int_{-\infty}^{\infty} W(\alpha)^4\, d\alpha < \infty$.

This assumption includes, for example, the quadratic spectral window or the Parzen window, but excludes other commonly used kernels such as the Daniell or Bartlett window (e.g., Priestley 1981).

With these assumptions, it can be shown that Q_T is appropriately standardized and under the null hypothesis has asymptotically a standard normal distribution.

Theorem 5 *Suppose that Assumptions 1–4 hold. Let $M_T \to \infty$ and $M_T^2/T \to 0$. If $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated, then $Q_T \to \mathcal{N}(0, 1)$ in distribution.*

In order to investigate the power of the proposed test for non-correlation, we consider sequences of local alternatives. More precisely, suppose that for $T \in \mathbb{N}$ the time series $\left(X_{t,1}^{(T)}\right)$ and $\left(X_{t,2}^{(T)}\right)$ satisfy Assumptions 1 and 2 and have cross-spectrum

$$f_{12}^{(T)}(\lambda) = c_T\, g(\lambda) \tag{4}$$

for some complex-valued function $g(\lambda)$. Then, if $c_T \to 0$ as T tends to infinity, the series $\left(X_{t,1}^{(T)}\right)$ and $\left(X_{t,2}^{(T)}\right)$ belong to a class of local alternatives that converge to the null hypothesis $H_0 : R_{12}(\lambda) = 0, \lambda \in [-\pi, \pi]$. If c_T decreases fast enough, the test statistic is still asymptotically normally distributed.

Theorem 6 *Suppose that, for $T \in \mathbb{N}$, the time series $\left(X_{t,1}^{(T)}\right)$ and $\left(X_{t,2}^{(T)}\right)$ satisfy Assumptions 1–4 and their cross-spectrum $f_{12}^{(T)}(\lambda)$ is of the form* (4). *Furthermore, let $M_T \to \infty$ and $M_T^2/T \to 0$ as $T \to \infty$ and let $c_T = M_T^{1/4}/T^{1/2}$. Then the test statistic Q_T is asymptotically normally distributed with mean*

$$\nu(g) = \frac{1}{\sigma} \int_{-\pi}^{\pi} \frac{|g(\lambda)|^2}{f_{11}(\lambda)\, f_{22}(\lambda)}\, d\lambda$$

and variance 1.

The asymptotic mean $\nu(g)$ under the sequence of local alternatives (4) has been called efficiacy of Q_T (e.g., Taniguchi et al. 1996) and measures the ability of the test to detect this sequence of local alternatives. We note that a similar result has been established for the time domain test statistic Q_T^* (Hong 1996, Theorem 2). Thus, the power of the two tests Q_T and Q_T^* can be compared by the asymptotic relative efficiency

$$\mathrm{ARE}\left(Q_T, Q_T^*\right) = \left(\frac{\mathrm{eff}(Q_T)}{\mathrm{eff}(Q_T^*)}\right)^2 = \left(\frac{\nu(g)}{\nu^*(g)}\right)^2,$$

where $\nu^*(g)$ is the asymptotic mean of Q_T^* under the sequence of local alternatives. Inserting the expressions for $\nu(g)$ and $\nu^*(g)$, we obtain

$$\mathrm{ARE}\left(Q_T, Q_T^*\right) = \frac{\sigma^{*2}}{\sigma^2} = \frac{H_2^4}{H_4^2} \leq 1,$$

where equality holds if and only if h is the rectangular taper function, which corresponds to the nontapered case. Thus, we find that the use of data taper leads to asymptotically less powerful tests. In the next section, we will show by simulations that tapering nevertheless is important and improves the small sample properties of the test statistic especially for time series that exhibit strong periodic behaviour.

4 Simulations

In this section, we present the results of two simulation studies that have been used to investigate the small sample properties of the proposed frequency domain test for non-correlation in comparison with the time domain version by Hong (1996). For the first simulation study, we consider a bivariate ARMA(1,1) model with processes $(X_{t,1})$ and $(X_{t,2})$ given by

$$\begin{aligned} X_{t,1} &= \tfrac{1}{2} X_{t-1,1} + \phi\, X_{t-1,2} + \varepsilon_{t,1}, \\ X_{t,2} &= \tfrac{1}{2} \varepsilon_{t-1,2} + \varepsilon_{t,2}. \end{aligned} \qquad (5)$$

For $\phi = 0$, the two series $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated and hence satisfy the null hypothesis.

For this model, samples of size $T = 100$ and 200 were generated both under the null hypothesis ($\phi = 0$) and under the alternative with $\phi = 0.2$. For each sample, the two test statistics Q_T and Q_T^* were calculated. The frequency domain test statistic Q_T was computed from the nontapered periodogram as well as from a tapered periodogram using a 20% cosine taper. For the computation of the time domain statistic Q_T^*, the two series were prewhitened using an AR(p) model with $p = 2, 4$, and 8. Additionally, the effect of the choice of bandwidth $B_T = 1/M_T$ and kernel w was examined by calculating all test statistics for six different bandwidths and two kernels, namely the rectangular kernel (or Daniell window) and the quadratic kernel (or Bartlett–Priestley window). While the latter satisfies Assumption 4, the former violates this assumption.

Table 1 reports the performance of the two tests under the null hypothesis at a 5% significance level, based on 10,000 replications. Here, the size of all test statistics depends on the bandwidth and decreases for smaller bandwidths. For fixed

Table 1 Rejection rates out of 10,000 replications with sample sizes $T = 100$ and 200 generated by model (5) under the null hypothesis ($\phi = 0.0$) for 5% significance level

		Frequency domain				Time domain					
		$h = 0\%$		$h = 20\%$		$p = 2$		$p = 4$		$p = 8$	
T	B_T	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*
100	0.3	3.7	6.9	2.2	4.8	4.8	5.0	4.9	5.1	5.8	6.1
	0.4	4.4	5.3	3.2	4.0	5.8	5.9	5.6	5.7	6.7	6.8
	0.5	5.1	7.4	4.4	6.8	6.1	6.4	6.4	6.3	7.6	7.5
	0.6	5.7	7.3	4.7	5.9	6.1	6.0	6.7	6.5	8.5	8.4
	0.7	6.5	7.3	5.5	6.0	6.4	6.2	7.0	6.8	8.2	8.2
	0.8	6.1	7.9	5.6	7.4	6.5	6.5	6.9	6.8	8.2	8.0
200	0.3	4.8	6.0	4.4	5.5	5.9	6.1	5.3	5.5	6.4	6.5
	0.4	5.8	7.5	5.2	6.8	6.5	6.7	6.4	6.5	7.1	6.9
	0.5	6.5	8.1	5.3	7.1	7.2	7.0	6.8	6.9	7.4	7.1
	0.6	6.3	7.0	6.0	6.4	6.6	6.4	6.5	6.6	7.4	7.4
	0.7	6.0	7.2	6.0	6.8	6.6	6.4	6.9	6.6	7.3	7.3
	0.8	6.3	7.7	6.8	8.0	7.1	6.9	6.9	7.1	7.5	7.5

QS quadratic-spectral kernel, *DN* Daniell window

bandwidth, the tapered version of Q_T yields smaller sizes than the nontapered version and hence requires more smoothing to obtain similar sizes. For Q_T^*, on the other hand, the rejection rate increases with the model order p. Furthermore, we note that the choice of kernel seems to affect the frequency domain test more than the time domain version: in the time domain, both kernels yield similar sizes for fixed order and bandwidth. Finally, comparing the results for the different sample sizes, we find that the size of the tests increases slightly with the sample size, which can be compensated by reducing the bandwidth accordingly. Summarizing we find that all tests have reasonable sizes although the frequency domain tests seem to be more susceptible to the choice of bandwidth.

Table 2 gives the power performances of the two tests under the alternative with $\phi = 0.2$ at a 5% significance level, again based on 10,000 replications. We note that the frequency domain tests show a large variation in power depending on bandwidth as well as on the choice of kernel or the amount of tapering. For fixed bandwidth, the nontapered version with rectangular kernel has highest power whereas the tapered version with quadratic kernel is the least powerful among all tests. In contrast, the time domain tests show only moderate changes in power as the bandwidth or the order p varies. These differences between the tests become less pronounced when comparing tests that have the same size under the null hypothesis although the results still show a loss in power when tapering is used.

In a second simulation study, we consider a bivariate ARMA(12,10) model with roots close to the unit circle (Table 3). The process thus exhibits strong periodic behaviour and the strong moving average part makes it difficult to prewhiten the process. The used model is

$$\begin{aligned} X_{t,1} &= \textstyle\sum_{u=1}^{12} \phi_u^{(1)}\, X_{t-u,1} + \phi^*\, X_{t-1,2} + \sum_{u=1}^{4} \theta_u^{(1)}\, \varepsilon_{t-u,1} + \varepsilon_{t,1}, \\ X_{t,2} &= \textstyle\sum_{u=1}^{4} \phi_u^{(2)}\, X_{t-u,2} + \sum_{u=1}^{12} \theta_u^{(2)}\, \varepsilon_{t-u,2} + \varepsilon_{t,2}. \end{aligned} \tag{6}$$

The coefficients that were used for the simulations are given in Table 3. For $\phi^* = 0$ the two series $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated and hence satisfy the null hypothesis.

Table 2 Rejection rates out of 10,000 replications with sample size $T = 100$ and 200 generated by model (5) under the alternative ($\phi = 0.2$) for 5% significance level

		Frequency domain				Time domain					
		$h = 0\%$		$h = 20\%$		$p = 2$		$p = 4$		$p = 8$	
T	B_T	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*
100	0.3	20.3	31.4	14.3	24.1	27.1	25.3	25.8	23.7	25.3	23.3
	0.4	25.5	31.0	20.8	25.8	30.8	28.9	30.0	27.5	29.4	27.7
	0.5	30.4	39.5	25.3	33.3	34.3	32.1	32.4	30.6	32.1	30.7
	0.6	34.2	39.9	28.0	33.1	35.8	34.4	34.0	32.9	34.1	33.0
	0.7	36.0	39.5	31.7	34.2	36.5	36.6	34.5	34.6	34.2	34.0
	0.8	37.0	42.5	32.2	37.0	36.2	36.7	35.2	35.6	35.3	35.1
200	0.3	52.2	59.5	44.0	51.7	58.4	54.7	58.8	54.7	56.5	52.4
	0.4	58.9	66.4	53.2	60.4	63.1	59.8	61.8	58.2	61.8	58.2
	0.5	64.4	70.8	56.7	63.9	67.3	64.6	66.3	63.7	64.4	61.9
	0.6	66.8	70.1	60.0	63.9	68.6	67.2	66.8	65.0	65.3	63.5
	0.7	69.6	72.1	63.9	67.5	69.3	68.4	67.3	67.1	66.0	65.4
	0.8	71.0	74.2	64.5	68.2	68.7	69.2	67.5	68.2	66.1	67.0

QS quadratic-spectral kernel, *DN* Daniell window

Table 3 Coefficients of bivariate ARMA(12,10) model (6)

u	1	2	3	4	5	6	7	8	9	10	11	12
$\phi_u^{(1)}$	4.7	−12.1	21.8	−30.8	35.9	−35.7	30.4	−22.2	13.4	−6.4	2.2	−0.4
$\phi_u^{(2)}$	2.3	−2.7	2.1	−0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\theta_u^{(1)}$	−0.1	0.1	−0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\theta_u^{(2)}$	2.6	4.1	4.8	4.6	4.2	4.2	3.9	2.9	1.6	0.5	0.0	0.0

Table 4 Rejection rates out of 10,000 replications with sample size $T = 1{,}000$ generated by model (6) under the null hypothesis ($\phi^* = 0.0$) for 5% significance level

	Frequency domain				Time domain							
	$h = 0\%$		$h = 20\%$		$p = 5$		$p = 10$		$p = 20$		$p = 40$	
B_T	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*
0.05	100.0	100.0	2.5	13.9	63.7	68.0	8.6	9.8	8.3	9.7	6.4	7.7
0.10	100.0	100.0	4.9	26.8	51.8	55.1	9.6	10.3	8.7	9.7	8.0	8.6
0.15	100.0	100.0	7.7	39.1	46.9	49.0	9.2	10.0	8.6	9.4	7.9	8.3
0.20	100.0	100.0	9.3	46.5	41.4	44.0	9.1	9.7	8.3	8.6	7.7	7.9

QS quadratic-spectral kernel, *DN* Daniell window

For the simulations, we used sample size $T = 1{,}000$. Table 4 reports the performances of the test statistics under the null hypothesis at a 5% significance level, based on 10,000 replications. The results show severe over-rejection for several test statistics. In particular, the nontapered version of Q_T breaks down completely and always rejects the null hypothesis of non-correlation: without tapering the periodogram estimates are strongly affected by leakage effects and thus may lead to severely biased estimates of the auto- and cross-spectra and, consequently, of the spectral coherence. Similarly, the time domain test leads to rejection rates of 40–70% if the autoregressive order is chosen too small ($p = 5$) to remove the serial correlation of the two series $(X_{t,1})$ and $(X_{t,2})$ sufficiently. For larger values

Table 5 Rejection rates out of 10,000 replications with sample size $T = 1,000$ generated by model (6) under the alternative ($\phi^* = 0.002$) for 5% significance level

	Frequency domain		Time domain							
	$h = 20\%$		$p = 5$		$p = 10$		$p = 20$		$p = 40$	
B_T	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*	*QS*	*DN*
0.05	30.2	66.5	96.0	96.3	38.6	37.7	45.4	45.2	42.8	42.4
0.10	57.6	89.1	93.4	94.4	45.1	46.5	51.3	53.0	51.8	53.3
0.15	75.0	96.1	89.5	92.2	42.5	46.9	48.5	53.0	48.5	52.6
0.20	83.7	98.3	84.0	88.8	38.0	44.0	43.6	49.6	43.7	49.9

QS quadratic-spectral kernel, *DN* Daniell window

of p, the tests still show moderate over-rejection, which decreases as p increases. Finally, we note that the use of a Daniell window for the frequency domain test yields rejection rates between 10 and 50%. These observations indicate that the smoothness of the kernel as required by Assumption 4 is indeed of importance. In this study, the nominal significance level of 5% is only attained by the frequency domain test based on a tapered periodogram and the quadratic-spectral window.

The results obtained under the alternative with $\phi^* = 0.002$ are given in Table 5. Here, the frequency domain test based on a tapered periodogram and the quadratic-spectral window shows rejection rates of 55–75% if we restrict ourselves to bandwidths that lead to reasonable sizes under the null hypothesis. In contrast, the time domain tests reject only in 35–55% of all cases. The results suggest that for processes with strong spectral features, the frequency domain test based on a tapered periodogram and a smooth kernel outperforms the other tests.

5 Bandwidth selection

One crucial step in the application of the proposed frequency domain test for non-correlation is the selection of an appropriate bandwidth for smoothing the periodogram. As we have seen in the previous section, the size and the power of the test both depend on the bandwidth and are sensitive to under- as well as to over-smoothing. Consequently, the application of the test requires a data-driven method for choosing the optimal bandwidth.

In the literature on nonparametric spectral density estimation, a number of criteria for bandwidth selection have been proposed; a partial overview and comparison is given, for example, in Fortin and Kuzmics (2000). In the following, we will consider three methods: the cross-validated log-likelihood (CVLL) criterion by Beltrão and Bloomfield (1987), a global version of the iterative procedure (ITP) suggested by Bühlmann (1996), and a method by Lee (2001) that combines plug-in and unbiased risk estimation (PURE) ideas. For all simulations, the frequency domain statistics were computed using a 20% cosine taper and a quadratic kernel for smoothing the periodogram.

Table 6 compares the empirical size and power of the frequency domain test for the three bandwidth selection methods when applied to samples of length $T = 100$, 200, and 300 from model (5). For sample size $T = 100$, the iterative method failed to converge in most cases and thus could not be used for testing. Otherwise, all

Table 6 Testing with data-driven bandwidth selection: rejection rates out of 5,000 replications for sample sizes $T = 100$, 200, and 500 generated by model (5) for 5% significance level

	$\phi = 0.0$			$\phi = 0.2$		
	$T = 100$	$T = 200$	$T = 500$	$T = 100$	$T = 200$	$T = 500$
PURE	6.6	6.2	6.5	32.5	61.7	95.7
CVLL	5.7	5.9	5.8	30.7	60.3	95.6
ITP	–	6.5	6.5	–	65.8	97.4

Table 7 Testing with data-driven bandwidth selection: rejection rates out of 5,000 replications for sample sizes $T = 500$, 1,000, and 2,000 generated by model (6) for testing for non-correlation at 5% significance level

	$\phi^* = 0.0$			$\phi^* = 0.002$		
	$T = 500$	$T = 1{,}000$	$T = 2{,}000$	$T = 500$	$T = 1{,}000$	$T = 2{,}000$
PURE	3.2	4.1	4.3	21.6	46.0	82.6
CVLL	0.8	3.0	3.9	6.0	30.8	80.7
ITP	6.3	6.1	6.3	38.0	69.5	93.8

three methods show good performance both under the null hypothesis and under the alternative and agree with the results in Tables 1 and 2. We note that the CVLL criterion achieves rejection frequencies that are closest to the nominal 5%-level; on the other hand, it has also the lowest power whereas the iterative method has the highest power of these tests and performs only slightly worse than the time domain tests (cf Table. 2).

In order to compare the merits of the three bandwidth selection criteria for hypothesis testing in less favourable situations, a second simulation study was performed with data generated from model (6). The results for samples of size T=500, 1,000, and 2,000 are given in Table 7. Here, the CVLL criterion appears to be severely biased for sample size $T = 500$ with rejection rate of 0.8% under the null hypothesis and 6% under the alternative. For larger sample sizes, the bias becomes less pronounced, but the test based on the CVLL criterion still performs worst. This observation is in accordance with the results by Fortin and Kuzmics (2000), who reported that the CVLL criterion has problems with sharp peak density processes. The best results are obtained by the iterative procedure of Bühlmann: while the test over-rejects slightly under the null hypothesis, it is uniformly and quite clearly more powerful than the tests based on the other two criteria. Moreover, we note that it clearly outperforms also the time domain tests: firstly, the time domain tests showed a higher tendency to over-reject under the null hypothesis and, secondly, they were far less powerful with rejection rates less than 55% as compared to 70% achieved by the frequency domain test with bandwidth selected by the ITP method.

6 Restriction to frequency bands

In some applications, one is interested in whether two time series $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated over a specific frequency interval. Suppose that $\Lambda = [\lambda_1, \lambda_2] \subseteq [0, \pi]$ is the frequency band of interest. Then integrated spectral coherence can be adapted to test for non-correlation over this frequency band by considering the statistic

$$S_{T,\Lambda} = \frac{1}{|J|} \sum_{j \in J} \left| \hat{R}_{12}^{(T)} \left(\frac{2\pi\, j}{T} \right) \right|^2,$$

where J is the set of all j such that $2\pi\, j/T \in \Lambda$. The corresponding standardized test statistic is

$$Q_{T,\Lambda} = \frac{T\, S_{T,\Lambda} - M_T\, \mu}{\sqrt{M_T/|\Lambda|}\, \sigma},$$

where $|\Lambda| = (\lambda_2 - \lambda_1)/\pi$ and μ and σ^2 are as in (2). We obtain the following result.

Theorem 7 *Suppose that the assumptions of Theorem 5 hold. Then under the null hypothesis*

$$R_{12}(\lambda) = 0 \quad \text{for}\, \lambda \in [\lambda_1, \lambda_2],$$

we have $Q_{T,\Lambda} \to \mathcal{N}(0, 1)$ *in distribution. Furthermore, under the sequence of local alternatives in* (4) *with* $c_T = M_T^{1/4}/T^{1/2}$, *we have* $Q_{T,\Lambda} \to \mathcal{N}(\nu_\Lambda(g), 1)$ *in distribution, where* $\nu_\Lambda(g)$ *is given by*

$$\nu_\Lambda(g) = \frac{1}{\sqrt{|\Lambda|}\, \sigma} \int_{\lambda_1}^{\lambda_2} \frac{|g(\lambda)|^2}{f_{11}(\lambda)\, f_{22}(\lambda)}\, d\lambda.$$

In order to compare the power of the localized test $Q_{T,\Lambda}$ and the global test Q_T, we consider the asymptotic relative efficiency

$$\text{ARE}\left(Q_T, Q_{T,\Lambda}\right) = \left(\frac{\nu(g)}{\nu_\Lambda(g)} \right)^2.$$

Suppose that the function g in (4) vanishes outside the interval Λ. From Theorems 6 and 7, we obtain

$$\text{ARE}\left(Q_T, Q_{T,\Lambda}\right) = |\Lambda| = \frac{\lambda_2 - \lambda_1}{\pi} \le 1,$$

that is, the global test Q_T is less powerful in detecting dependencies restricted to the frequency range $[\lambda_1, \lambda_2]$ than the localized test $Q_{T,\Lambda}$. This is due to the fact that the global test Q_T tests against a larger class of alternatives than the localized test.

For an illustration of the advantages of the localized test for non-correlation, we consider neurological data concerned with the identification of signal transmission pathways in the investigation of human tremor. Tremor is defined as the involuntary, oscillatory movement of parts of the body, mainly the upper limbs. It has been shown that patients suffering from Parkinson's disease show a tremor-related cortical activity which can be detected in the EEG time series by cross-spectral analysis (e.g., Timmer et al. 2000).

Table 8 Results of testing for non-correlation between EEG channels C2P and C3 and EMG signal of left hand wrist extensor

	C2P-EMG	C3-EMG	C2P-C3
Q_T	0.97	−0.34	3.81
$Q^*_{T,p=1,000}$	0.54	0.01	2.89
$Q_{T,[1\,\text{Hz},8\,\text{Hz}]}$	12.16	0.08	0.35

In one experiment, the EEG and the surface electromyogram (EMG) of the left hand wrist extensor muscle in a healthy subject have been measured during an externally enforced oscillation of the hand with 1.9 Hz. All data were simultaneously sampled at 1,000 Hz. For the analysis, a subseries of length T=40,000 has been used. All kernel estimates were computed with bandwidth $B_T = 0.002$. The data are described in detail in Dahlhaus and Eichler (2003). In the following, we restrict ourselves to two EEG channels, namely C2P and C3.

Figure 1a displays estimates for the spectral densities of the three series. The spectrum of the EMG signal shows a strong periodicity of about 1.9 Hz. The higher harmonics in the spectrum indicate that the process has a nonlinear dynamic. The signal for channel C2P also shows strong peaks at the oscillation frequency of 1.9 Hz and the first harmonic, whereas the signal for channel C3 does not show any clear signs of periodicity.

The estimates for the spectral coherences between the three time series in Figure 1b indicate a strong relationship between C2P and the EMG signal at the oscillation frequency and the first harmonic, while for the other two pairs the spectral coherence exhibits only a small peak at the first harmonic. For a statistical evaluation of the relationships between the three series, we tested for non-correlation between each pair of series. Because of the strong periodicity and nonlinearity of the EMG signal, the autoregressive model order used for prewhitening the series in the computation of the time domain test was very large (p=1,000). For the frequency domain test, we used a tapered periodogram with 10% cosine taper. The values of the test statistics Q_T, Q^*_T, and $Q_{T,[1\text{Hz},8\text{Hz}]}$ are given in Table 8. Here, both global tests for non-correlation do not find a significant relationship between the EMG signal and either EEG channel, but establish a significant relationship between the two EEG channels.

On the other hand, if we test for non-correlation only over the frequency band from 1 to 8 Hz, which is most relevant for the discussion of tremor-like phenomena, channel C2P and the EMG signal show a highly significant relationship, whereas neither C3 and the EMG signal nor the two EEG channels are significantly related over this frequency band. We conclude that only channel C2P, which is located over the right hemisphere, shows a cortical activity that is directly related with the oscillating movement of the left hand.

7 Conclusions

We have presented a new frequency domain based test for non-correlation between two stationary time series. The test is based on the integrated spectral coherence and has similar asymptotic properties as a previously introduced time domain

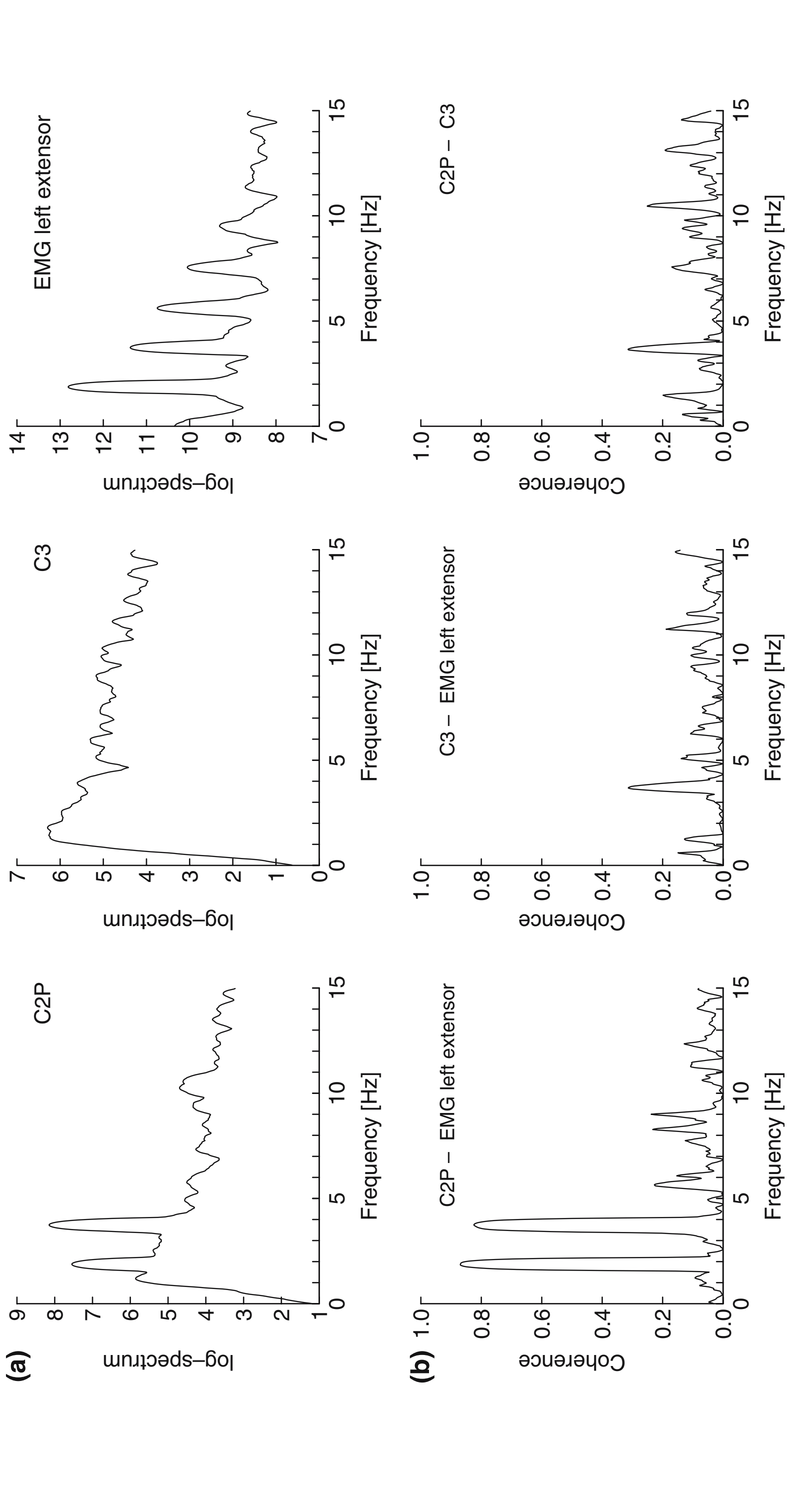

Fig. 1 **a** Spectra for EEG channels C2P and C3 and for EMG from left wrist extensor. **b** Coherencies between EEG channels C2P and C3 and EMG from left wrist extensor

version. Unlike the time domain version, the proposed test is truly nonparametric and does not require prewhitening of the time series by autoregressive model fitting. Consequently, the test shows a high power of detecting relationships also for highly periodic, nonlinear processes, for which prewhitening by autoregressive modelling is less appropriate.

The performance of the test crucially depends on the chosen bandwidth for smoothing the periodogram. Thus, a data-driven method for selecting the optimal bandwidth is needed to make the proposed test applicable. We have compared three bandwidth selection methods for spectral density estimation. Our results indicate that a global version of the iterative bandwidth selection method by Bühlmann (1996) achieves good results, although it can suffer from convergence problems for small sample sizes. In that case, it can be replaced, for example, by the PURE bandwidth selection method of Lee (2001).

Moreover, we have shown that a modification of the test can be used to test for non-correlation over frequency bands. The restriction to frequency bands of interest allows a more powerful detection of relationships at those frequencies.

Acknowledgements The author wishes to thank two anonymous referees for their comments and remarks, which helped to improve on an earlier version.

Appendix: Proofs

The derivation of the asymptotic distribution of the test statistic Q_T is based on the following version of the integrated spectral coherence. Let

$$\tilde{S}_T = \frac{1}{2\pi} \int_{\Pi} \left| \hat{R}_{12}^{(T)}(\lambda) \right|^2 d\lambda$$

with $\Pi = [-\pi, \pi]$, where the coherence $\hat{R}_{12}^{(T)}(\lambda)$ is computed from spectral estimates

$$\tilde{f}_{ab}^{(T)}(\lambda) = \int_{\Pi} w^{(T)}(\lambda - \alpha)\, I_{ab}^{(T)}(\alpha)\, d\alpha.$$

Using a Taylor expansion of second order, $\tilde{S}_T$ can be approximated by

$$\tilde{S}_T^{(2)} = \int_{\Pi} \frac{\tilde{f}_{12}^{(T)}(\lambda)\, \tilde{f}_{21}^{(T)}(\lambda)}{2\pi\, f_{11}(\lambda)\, f_{22}(\lambda)}\, d\lambda$$

with $\tilde{S}_T = \tilde{S}_T^{(2)} + o_P\left(M_T^{1/2}/T\right)$.

For the proofs, we need the following function, which has been introduced by Dahlhaus (1983). For $T > 0$ let $L^{(T)}$ be the 2π-periodic function with

$$L^{(T)}(\lambda) = \begin{cases} T, & |\lambda| \leq 1/T \\ \frac{1}{|\lambda|}, & 1/T < |\lambda| \leq \pi. \end{cases} \tag{7}$$

The properties of these functions are summarized by the following lemma. Here and throughout the paper, C denotes a generic constant.

Lemma 1 *Let $L^{(T)}(\lambda)$ be defined as in* (7), $\alpha, \beta, \gamma \in \mathbb{R}$ *and* $r \in \mathbb{N}$. *We obtain with a constant C independent of T, T_1, and T_2*

(i) $L^{(T)}(\alpha)$ *is monotonically increasing in* $T \in \mathbb{R}^+$ *and decreasing in* $\alpha \in [0, \pi]$.
(ii) $L^{(T)}(c\alpha) \le c^{-1} L^{(T)}(\alpha)$ *for all* $c \in (0, 1]$.

(iii) $$\int_{\Pi} L^{(T)}(\alpha) d\alpha \le C \log(T).$$

(iv) $$\int_{\Pi} L^{(T_1)}(\beta+\alpha) L^{(T_2)}(\gamma-\alpha) d\alpha \le C \max\{\log(T_1), \log(T_2)\} L^{(\min\{T_1, T_2\})}(\beta+\gamma).$$

(v) $$\int_{\Pi} L^{(T)}(\alpha)^r d\alpha \le C T^{r-1}.$$

(vi) $$\int_{\Pi} L^{(T_1)}(\beta+\alpha)^r L^{(T_2)}(\gamma-\alpha)^r d\alpha \le C \max\{T_1^{r-1}, T_2^{r-1}\} L^{(\min\{T_1, T_2\})}(\beta+\gamma)^r.$$

Proof The proofs are straightforward and can be found in Dahlhaus (1983, 1990).

The $L^{(T)}$ function provides convenient upper bounds when dealing with kernels and data tapers in the frequency domain. More precisely, suppose that h is a taper function of bounded variation and let $H_{k,T}(\lambda) = \sum_{t=1}^{T} h_{t,T}^k \exp(-\mathrm{i}\lambda t)$ the corresponding finite Fourier transform. Then, we have

$$\left| H_{k,T}(\lambda) \right| \le C\, L^{(T)}(\lambda) \tag{8}$$

with positive constant C independent of T and λ. Similarly, we obtain by Assumption 4 for the kernel $w^{(T)}$

$$w^{(T)}(\lambda) \le C \frac{L^{(M_T)}(\lambda)^2}{M_T},$$

again with a positive constant C independent of T and λ.

For the derivation of the asymptotic distribution of $\tilde{S}_T$, we follow the approach due to Brillinger (1981) and show convergence of the cumulants of first, second, and higher order to the corresponding cumulants of the limit distribution in Lemmas 4, 6, and 7, respectively. This leads us to considering cumulants of the form

$$\operatorname{cum}\left\{ d_{i_{j,1}}^{(T)}(\alpha_{j,1}) d_{i_{j,2}}^{(T)}(-\alpha_{j,1}) d_{i_{j,3}}^{(T)}(\alpha_{j,2}) d_{i_{j,4}}^{(T)}(-\alpha_{j,2}) \mid j = 1, \ldots, k \right\}. \tag{9}$$

In the following, we introduce the basic ideas and concepts for evaluating these cumulants. First, we note that by the product theorem for cumulants (cf Brillinger 1981, Theorem 2.3.2), the above cumulant is equal to

$$\sum_{i.p.} \prod_{j=1}^{m} \operatorname{cum}\left\{ d_{\nu_{j,1}}^{(T)}(\gamma_{j,1}), \ldots, d_{\nu_{j,p_j}}^{(T)}(\gamma_{j,p_j}) \right\},$$

where $\sum_{i.p.}$ denotes the sum over all indecomposable partitions $\{P_1, \ldots, P_m\}$ of the table

$$\begin{matrix} \alpha_{1,1} & -\alpha_{1,1} & \alpha_{1,2} & -\alpha_{1,2} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{k,1} & -\alpha_{k,1} & \alpha_{k,2} & -\alpha_{k,2} \end{matrix} \tag{10}$$

with $p_j = |P_j|$ and $P_j = \{\gamma_{j,1}, \ldots, \gamma_{j,p_j}\}$. Recall that a partition $\{P_1, \ldots, P_m\}$ is said to be indecomposable if every set P_j is hooked to at least one other set P_i, where two sets P_i and P_j are hooked if there exists an index $l \in \{1, \ldots, k\}$ and variables $\gamma_{i_r} \in P_i$ and $\gamma_{i_{r'}} \in P_j$ such that γ_{i_r} and $\gamma_{i_{r'}}$ are both contained in the lth row $\{\alpha_{l,1}, -\alpha_{l,1}, \alpha_{l,2}, -\alpha_{l,2}\}$.

Next, we note that

$$\begin{aligned} &\operatorname{cum}\left\{d_{a_1}^{(T)}(\alpha_1), \ldots, d_{a_k}^{(T)}(\alpha_k)\right\} \\ &\quad = (2\pi)^{k-1} H_{k,T}(\alpha_1 + \cdots + \alpha_k) f_{a_1\cdots a_k}(\alpha_1, \ldots, \alpha_{k-1}) + O(1) \end{aligned} \tag{11}$$

uniformly in $\alpha_1, \ldots, \alpha_k \in \Pi$ (Brillinger 1981, Theorem 4.3.2), where

$$\begin{aligned} &f_{a_1\cdots a_k}(\alpha_1, \ldots, \alpha_{k-1}) = (2\pi)^{-k+1} \\ &\quad \textstyle\sum_{u_1,\ldots,u_{k-1}\in\mathbb{Z}} c_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1}) \exp\left(-\mathrm{i}(u_1\alpha_1 + \cdots + u_{k-1}\alpha_{k-1})\right) \end{aligned}$$

is the kth order cumulant spectrum of the process. Thus, we obtain for the cumulants in (9)

$$\begin{aligned} &\operatorname{cum}\left\{d_{i_{j,1}}^{(T)}(\alpha_{j,1}) d_{i_{j,2}}^{(T)}(-\alpha_{j,1}) d_{i_{j,3}}^{(T)}(\alpha_{j,2}) d_{i_{j,4}}^{(T)}(-\alpha_{j,2}) \mid j = 1, \ldots, k\right\} \\ &\quad = \sum_{i.p.} \prod_{j=1}^{m} \left[(2\pi)^{p_j-1} H_{p_j,T}(\bar{\gamma}_j) f_{\nu_{j,1}\cdots\nu_{j,p_j}}(\gamma_{j,1}, \ldots, \gamma_{j,p_j-1}) + O(1)\right] \\ &\quad = \sum_{i.p.} \prod_{j=1}^{m} (2\pi)^{p_j-1} H_{p_j,T}(\bar{\gamma}_j) f_{\nu_{j,1}\cdots\nu_{j,p_j}}(\gamma_{j,1}, \ldots, \gamma_{j,p_j-1}) + R_T, \end{aligned} \tag{12}$$

where $\bar{\gamma}_j = \gamma_{j,1} + \cdots + \gamma_{j,p_j}$ and the remainder term R_T satisfies

$$|R_T| \le C \sum_{i.p.} \sum_{J \subsetneq \{1,\ldots,m\}} \prod_{j\in J} (2\pi)^{p_j-1} \left|H_{p_j,T}(\bar{\gamma}_j)\right| \left|f_{\nu_{j,1}\ldots\nu_{j,p_j}}(\gamma_{j,1}, \ldots, \gamma_{j,p_j-1})\right|.$$

Noting that, by Assumption 1, the kth order cumulant spectra are bounded, we can use (8) to obtain an upper bound for (12). The further evaluation of these bounds is based on Lemma 1 and the following lemma.

Lemma 2 *Let $k \ge 3$ and suppose that $\{P_1, \ldots, P_m\}$ is an indecomposable partition of the table*

$$\begin{matrix} \alpha_1 & -\alpha_1 \\ \vdots & \vdots \\ \alpha_k & -\alpha_k. \end{matrix}$$

If $m = k$, then for any $k-2$ variables $\alpha_{i_1}, \dots, \alpha_{i_{k-2}}$ we obtain

$$\int_{\Pi^{k-2}} \prod_{j=1}^{k} L^{(T)}(\bar{\gamma}_j) d\alpha_{i_1} \cdots d\alpha_{i_{k-2}} \leq C L^{(T)}(\alpha_{i_{k-1}} \pm \alpha_{i_k})^2 \log(T)^{k-2}.$$

If $m < k$, there exist $k-2$ variables $\alpha_{i_1}, \dots, \alpha_{i_{k-2}}$ such that

$$\int_{\Pi^{k-2}} \prod_{j=1}^{m} L^{(T)}(\bar{\gamma}_j) d\alpha_{i_1} \cdots d\alpha_{i_{k-2}} \leq C T \log(T)^{k-2}.$$

Proof The first part follows from the indecomposability of the partition and the properties of the $L^{(T)}$-function. For the second part, we note that because of the indecomposability of the partition there exists an ordering $P_{j_1}, \dots, P_{j_m}$ and variables $\alpha_{i_1}, \dots, \alpha_{i_{m-1}}$ such that $\alpha_{i_r} \in \bigcup_{i=1}^{r} P_{j_i}$ and $-\alpha_{i_r} \in P_{j_{r+1}}$. Therefore, we have

$$\begin{aligned} &\int_{\Pi^{k-2}} L^{(T)}(\bar{\gamma}_1) \cdots L^{(T)}(\bar{\gamma}_{m-1}) d\alpha_{i_1} \cdots d\alpha_{i_{k-2}} \\ &\leq C \log(T) \int_{\Pi^{k-3}} L^{(T)}(\bar{\gamma}_1 + \bar{\gamma}_2) \cdots L^{(T)}(\bar{\gamma}_{m-1}) d\alpha_{i_2} \cdots d\alpha_{i_{k-2}} \\ &\leq C \log(T)^{m-1} L^{(T)}(\bar{\gamma}_1 + \cdots + \bar{\gamma}_m) = C T \log(T)^{m-1}, \end{aligned} \tag{13}$$

which concludes the proof.

For the application of this lemma, it will be convenient to consider a partition $\{P_1, \dots, P_m\}$ of table (10) also as a partition of the table

$$\begin{array}{cc} \alpha_{1,1} & -\alpha_{1,1} \\ \alpha_{1,2} & -\alpha_{1,2} \\ \vdots & \vdots \\ \alpha_{k,1} & -\alpha_{k,1} \\ \alpha_{k,2} & -\alpha_{k,2} \end{array}. \tag{14}$$

This partition, however, might now be decomposable into two or more indecomposable subpartitions. We call these subpartitions non-hooked as two sets from different subpartitions are not hooked. Furthermore, we say that a subpartition covers only one variable if it consists of exactly one row of the table (14), that is, it consists either of one set $\{\alpha_{l,i}, -\alpha_{l,i}\}$ or of two sets $\{\alpha_{l,i}\}$ and $\{-\alpha_{l,i}\}$ for some $l \in \{1, \dots, k\}$ and $i \in \{1, 2\}$.

When applying the first part of Lemma 2 repeatedly to the non-hooked subpartitions of table (14), we need to choose $\alpha_{i_{k-1}}$ and α_{i_k} such that subsequent steps in the evaluation lead to the lowest possible bounds. To this end, we introduce the following concept of a circle.

For each partition $\{P_1, \ldots, P_m\}$, we denote by $Q_j = \{\nu_{j,1}, \ldots, \nu_{j,p_j}\}$ the sets of the corresponding partition of the table of indices

$$\begin{matrix} i_{1,1} & i_{1,2} & i_{1,3} & i_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ i_{k,1} & i_{k,2} & i_{k,3} & i_{k,4}. \end{matrix} \tag{15}$$

Suppose that $\{P_1, \ldots, P_m\}$ has s non-hooked subpartitions of table (14), of which u cover only one variable. Now suppose that we select for each of the $s-u$ remaining non-hooked subpartitions two variables $\alpha_{l_{j,1},i_{j,1}}$ and $\alpha_{l_{j,2},i_{j,2}}$ such that $l_{j,1} \neq l_{j,2}$ for all $j = 1, \ldots, s-u$. Then a sequence $(l_{j_1,1}, l_{j_1,2}), \ldots, (l_{j_r,1}, l_{j_r,2})$ in the set $\{(l_{j,1}, l_{j,2}) | j = 1, \ldots, s-u\}$ is called a circle if

$$l_{j_u,2} = l_{j_{u+1},1} \quad \forall u = 1, \ldots, r-1 \text{ and } l_{j_r,2} = l_{j_1,1}.$$

The next lemma states that the variables $\alpha_{l_{j,1},i_{j,1}}$ and $\alpha_{l_{j,2},i_{j,2}}$ can be chosen such that not more than one circle is obtained.

Lemma 3 *Let $J_1, \ldots, J_{s-u}$ denote the non-hooked subpartitions of table* (14) *that cover more than one variable. Then, for each subpartition J_r, there exist two variables $\alpha_{l_{r,1},i_{r,1}}$ and $\alpha_{l_{r,2},i_{r,2}}$ covered by J_r such that the set $\{(l_{j,1}, l_{j,2}) | j = 1, \ldots, s-u\}$ does not contain more than one circle.*

Proof Unifying the sets within each subpartition, we obtain again an indecomposable partition $Q_1, \ldots, Q_s$ of the table (10). Since the u subpartitions that cover only one variable do not link different rows they can be omitted without destroying indecomposability. Now, suppose there are two circles represented by the sets $Q_{j_1}, \ldots, Q_{j_r}$ and $Q_{j_{r+1}}, \ldots, Q_{j_{r'}}$, respectively. Then the sets $A = Q_{j_1} \cup \cdots \cup Q_{j_r}$, $B = Q_{j_{r+1}} \cup \cdots \cup Q_{j_{r'}}$, and $Q_{j_{r'+1}}, \ldots, Q_{j_{s-u}}$ form a new indecomposable partition. Therefore there exists a sequence $A, \tilde{Q}_1, \ldots, \tilde{Q}_q, B$ such that two consecutive sets are hooked. Choosing the variables correspondingly we obtain a selection with at least one circle less. As there can be only a finite number of circles, we can apply this scheme repeatedly.

We note that the lemma remains true if we consider only a subset of the non-hooked subpartitions. Finally, we define the sequence $\{\Phi_2^{(T)}\}_{T\in\mathbb{N}}$ of functions

$$\Phi_2^{(T)}(\lambda) = \frac{|H_{2,T}(\lambda)|^2}{2\pi \; H_{4,T}(0)}. \tag{16}$$

By the upper bound in (8) and Lemma 1, it can be shown that $\{\Phi_2^{(T)}\}_{T\in\mathbb{N}}$ is an approximate identity (e.g., Dahlhaus 1983).

Lemma 4 *Suppose that Assumptions 1–4 hold. Furthermore, let $M_T \to \infty$ and $M_T^2/T \to 0$. Then, if $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated, we have*

$$\mathbb{E}\left(\tilde{S}_T^{(2)}\right) = \frac{M_T}{T} \frac{2\pi \; H_4}{H_2^2} \int_{-\infty}^{\infty} w(\alpha)^2 \, d\alpha + o\left(\frac{M_T^{1/2}}{T}\right). \tag{17}$$

Proof Let $\Delta(\lambda) = (2\pi\, f_{11}(\lambda)\, f_{22}(\lambda))^{-1}$. Noting $\operatorname{cum}\{d_a^{(T)}(\alpha)\} = 0$, it then follows from the product theorem for cumulants

$$\mathbb{E}\left[\int_{\Pi} \Delta(\lambda)\, \tilde{f}_{12}^{(T)}(\lambda)\, \tilde{f}_{21}^{(T)}(\lambda)\, d\lambda\right]$$
$$= \frac{1}{(2\pi H_{2,T}(0))^2} \int_{\Pi^3} \Delta(\lambda)\, w^{(T)}(\lambda-\alpha)\, w^{(T)}(\lambda-\beta)$$
$$\times\Big[\operatorname{cum}\{d_1^{(T)}(\alpha), d_2^{(T)}(-\alpha), d_2^{(T)}(\beta), d_1^{(T)}(-\beta)\}$$
$$+\operatorname{cum}\{d_1^{(T)}(\alpha), d_2^{(T)}(-\alpha)\}\operatorname{cum}\{d_2^{(T)}(\beta), d_1^{(T)}(-\beta)\}$$
$$+\operatorname{cum}\{d_1^{(T)}(\alpha), d_2^{(T)}(\beta)\}\operatorname{cum}\{d_2^{(T)}(-\alpha), d_1^{(T)}(-\beta)\}$$
$$+\operatorname{cum}\{d_1^{(T)}(\alpha), d_1^{(T)}(-\beta)\}\operatorname{cum}\{d_2^{(T)}(-\alpha), d_2^{(T)}(\beta)\}\Big]\, d\alpha\, d\beta\, d\lambda. \quad (18)$$

Since $d_1^{(T)}(\lambda)$ and $d_2^{(T)}(\lambda)$ are uncorrelated, the second and third terms are zero. Furthermore, by (11), we find that the first term is of order $O(T^{-1})$ while the last term can be rewritten as

$$\frac{2\pi H_4}{T H_2^2} \int_{\Pi^3} \Delta(\lambda)\, f_{11}(\lambda-\alpha)\, f_{22}(\lambda-\alpha)\, w^{(T)}(\alpha)\, w^{(T)}(\alpha-\beta)\, \Phi_2^{(T)}(\beta)$$
$$d\alpha\, d\beta\, d\lambda + O(T^{-1}).$$

Let $\psi(\lambda, \alpha) = |\Delta(\lambda)\, f_{11}(\lambda-\alpha)\, f_{22}(\lambda-\alpha)|$. Since w and ψ are Lipschitz continuous, we obtain with $L^{(T)}(\lambda) \le |\lambda|^{-1}$

$$\int_{\Pi^3} \psi(\lambda, \alpha)\, |w^{(T)}(\alpha)\, w^{(T)}(\alpha-\beta) - w^{(T)}(\alpha)^2|\, \Phi_2^{(T)}(\beta)\, d\alpha\, d\beta\, d\lambda$$
$$\le C\, M_T^2 \int_{\Pi} |\beta|\, \Phi_2^{(T)}(\beta)\, d\beta \le \frac{C\, M_T^2}{T} \int_{\Pi} L^{(T)}(\beta)\, d\beta \le \frac{C\, M_T^2\, \log(T)}{T},$$

and, noting that $\int_{\Pi} \Phi_2^{(T)}(\beta)\, d\beta = 1$,

$$\int_{\Pi^3} |\psi(\lambda, \alpha) - \psi(\lambda, 0)|\, w^{(T)}(\alpha)^2\, \Phi_2^{(T)}(\beta)\, d\alpha\, d\beta\, d\lambda$$
$$\le \frac{C}{M_T^2} \int_{\Pi} |\lambda| L^{(M_T)}(\lambda)^4 d\lambda \le C.$$

Together with $M_T^{-1} \int_{\Pi} w^{(T)}(\alpha)^2\, d\alpha \to \int w(\alpha)^2\, d\alpha$, this proves the convergence in (17).

For the derivation of the variance of $\tilde{S}_T^{(2)}$, we define the sequence $\{\Psi^{(T)}\}_{T\in\mathbb{N}}$ of functions

$$\Psi^{(T)}(\alpha_1,\dots,\alpha_5)=\frac{1}{C_\Psi^{(T)}}w^{(T)}(\alpha_1)\cdots w^{(T)}(\alpha_4)\Phi_2^{(T)}(\alpha_5)\Phi_2^{(T)}(\alpha_1+\alpha_2-\alpha_3-\alpha_4+\alpha_5)$$

with

$$C_\Psi^{(T)}=\int_{\Pi^5} w^{(T)}(\alpha_1)\cdots w^{(T)}(\alpha_4)\Phi_2^{(T)}(\alpha_5)\Phi_2^{(T)}(\alpha_1+\alpha_2-\alpha_3-\alpha_4+\alpha_5)d\alpha_1\cdots d\alpha_5.$$

Lemma 5 *Let w satisfy Assumption 4. Then*

(i) $\Psi^{(T)}$ *is an approximate identity;*
(ii) $\lim_{T\to\infty}\frac{1}{M_T}C_\Psi^{(T)}=\frac{1}{2\pi}\int_{-\infty}^{\infty}W(\alpha)^4d\alpha;$
(iii) for any integrable function $g:\mathbb{R}^6\to\mathbb{R}$, *we have*

$$\Big|\int_{\Pi^6} g(\lambda,\alpha_1,\dots,\alpha_5)\,\Psi^{(T)}(\alpha_1,\dots,\alpha_5)\,d\alpha_1\cdots d\alpha_5\,d\lambda$$
$$-\int_{\Pi} g(\lambda,0,\dots,0)\,d\lambda\Big|=o(1).$$

Proof The proof is straightforward.

Lemma 6 *Suppose that Assumptions 1–4 hold. Furthermore, let $M_T\to\infty$ and $M_T^2/T\to 0$. Then, if $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated, we have*

$$\mathbb{E}\left(\tilde{S}_T^{(2)}\right)=\frac{M_T}{T^2}\frac{2\,H_4^2}{H_2^4}\int_{-\infty}^{\infty}W(\lambda)^4d\lambda+o\left(\frac{M_T}{T^2}\right).$$

Proof Let $\Delta(\lambda)$ be as in the proof of Lemma 4. From (12) we obtain for the variance of $\tilde{S}_T^{(2)}$

$$\begin{aligned}
&\mathrm{var}\left(\tilde{S}_T^{(2)}\right)\\
&=\int_{\Pi^6}\prod_{j=1}^{2}\Big[\Delta(\lambda_j)\,w^{(T)}(\lambda_j-\alpha_{j,1})\,w^{(T)}(\lambda_j-\alpha_{j,2})\Big]\\
&\quad\times\mathrm{cum}\{I_{12}^{(T)}(\alpha_{1,1})\,I_{21}^{(T)}(\alpha_{1,2}),\,I_{12}^{(T)}(\alpha_{2,1})\,I_{21}^{(T)}(\alpha_{2,2})\}\,d\alpha_{1,1}\cdots d\alpha_{2,2}\,d\lambda_1\,d\lambda_2\\
&=\frac{1}{(2\pi\,H_{2,T})^4}\sum_{i.p.}\int_{\Pi^6}\prod_{j=1}^{2}\Big[\Delta(\lambda_j)\,w^{(T)}(\lambda_j-\alpha_{j,1})\,w^{(T)}(\lambda_j-\alpha_{j,2})\Big]\\
&\quad\times\prod_{j=1}^{m}(2\pi)^{p_j-1}H_{p_j,T}(\bar{\gamma}_j)\,f_{\nu_{j,1}\dots\nu_{j,p_j}}(\gamma_{j,1},\dots,\gamma_{j,p_j-1})\,d\alpha_{1,1}\cdots\\
&\quad d\alpha_{2,2}\,d\lambda_1\,d\lambda_2+R_T,
\end{aligned}\tag{19}$$

where the remainder term R_T is of smaller order than the main term. We evaluate the terms for the different partitions separately. First, we consider the partition consisting of the sets $\{\alpha_{1,1}, \alpha_{2,1}\}$, $\{-\alpha_{1,1}, -\alpha_{2,1}\}$, $\{\alpha_{1,2}, \alpha_{2,2}\}$, and $\{-\alpha_{1,2}, -\alpha_{2,2}\}$. We obtain terms of the form

$$\frac{1}{H_2^4 T^4} \int\limits_{\Pi^6} \prod_{j=1}^{2} \Delta(\lambda_j)\, w^{(T)}(\lambda_j - \alpha_{j,1})\, w^{(T)}(\lambda_j - \alpha_{j,2}) \left| H_{2,T}(\alpha_{1,j} + \alpha_{2,j}) \right|^2$$
$$\times f_{11}(\alpha_{1,1})\, f_{22}(-\alpha_{1,1})\, f_{22}(\alpha_{1,2})\, f_{11}(-\alpha_{1,2})\, d\alpha_{1,1} \cdots d\alpha_{2,2}\, d\lambda_1\, d\lambda_2$$
$$= \frac{(2\pi)^2 C_\Psi^{(T)} H_4^2}{H_2^4 T^2} \int\limits_{\Pi^6} \Delta(\lambda_1)\Delta(\lambda_2 - \lambda_1 + \alpha_{1,1} + \alpha_{2,1})$$
$$\times f_{11}(\lambda_1 - \alpha_{1,1})\, f_{22}(\alpha_{1,1} - \lambda_1)\, f_{22}(\lambda_1 - \alpha_{1,2})\, f_{11}(\alpha_{1,2} - \lambda_1)$$
$$\times \Psi^{(T)}(\alpha_{1,1}, \alpha_{1,2}, \alpha_{2,1}, \alpha_{2,2}, \lambda_2)\, d\alpha_{1,1} \cdots d\alpha_{2,2}\, d\lambda_1\, d\lambda_2$$

and further by Lemma 5(iii)

$$= \frac{(2\pi)^2 C_\Psi^{(T)} H_4^2}{H_2^4 T^2} \int\limits_{\Pi} \Delta(\lambda)\, \Delta(\lambda)\, f_{11}(\lambda)\, f_{22}(\lambda)\, f_{11}(\lambda)\, f_{22}(\lambda)\, d\lambda + o\left(\frac{M_T}{T^2}\right).$$

Inserting the definition of $\Delta(\lambda)$ we obtain

$$\frac{M_T}{T^2} \frac{H_4^2}{H_2^4} \int\limits_{-\infty}^{\infty} W(\alpha)^4\, d\alpha + o\left(\frac{M_T}{T^2}\right).$$

For the partition $\{\alpha_{1,1}, -\alpha_{2,2}\}$, $\{-\alpha_{1,1}, \alpha_{2,2}\}$, $\{\alpha_{1,2}, -\alpha_{2,1}\}$, $\{-\alpha_{1,2}, \alpha_{2,1}\}$, we obtain the same result. In all other partitions of length $m = 4$, at least one factor is the cross-spectrum $f_{12}(\alpha)$ and, thus, the term is zero. The same argument applies to any partition with a set of the form $\{\alpha, -\alpha\}$.

Next, we show that for all other partitions the corresponding term in (19) is of order $o(M_T/T^2)$. First, if the partition consists of only one or two sets, we directly get the upper bound

$$\frac{C}{T^4} \int\limits_{\Pi^6} w^{(T)}(\lambda_1 - \alpha_{1,1}) \cdots w^{(T)}(\lambda_2 - \alpha_{2,2}) T^2 d\alpha_{1,1} \cdots d\alpha_{2,2} d\lambda_1 d\lambda_2 \leq \frac{C}{T^2}.$$

Second, if all sets in the partition are also hooked in table (10), we obtain

$$\frac{C}{T^4M_T^4}\int_{\Pi^6} L^{(M_T)}(\lambda_1-\alpha_{1,1})^2\cdots L^{(M_T)}(\lambda_2-\alpha_{2,2})^2$$
$$\times\prod_{j=1}^{m} L^{(T)}(\bar{\gamma}_j)d\alpha_{1,1}\cdots d\alpha_{2,2}d\lambda_1 d\lambda_2$$
$$\leq \frac{C}{T^4M_T^2}\int_{\Pi^4} L^{(M_T)}(\alpha_{1,1}-\alpha_{1,2})^2 L^{(M_T)}(\alpha_{2,1}-\alpha_{2,2})^2$$
$$\times\prod_{j=1}^{m} L^{(T)}(\bar{\gamma}_j)d\alpha_{1,1}\cdots d\alpha_{2,2}$$
$$\leq \frac{CM_T^2}{T^4}\int_{\Pi^4}\prod_{j=1}^{m} L^{(T)}(\bar{\gamma}_j)d\alpha_{1,1}\cdots d\alpha_{2,2}$$
$$\leq \frac{CM_T^2\log(T)^{m-2}}{T^3}=o\left(\frac{M_T}{T^2}\right).$$

Finally, we consider partitions of length $m=3$ with two non-hooked subpartitions. If one set, P_1 say, spans two variables, the other two sets are hooked and have two elements each. It follows that $\bar{\gamma}_2=-\bar{\gamma}_3$ and, therefore, we obtain

$$\frac{C}{T^4M_T^4}\int_{\Pi^6} L^{(M_T)}(\lambda_1-\alpha_{1,1})^2\cdots L^{(M_T)}(\lambda_2-\alpha_{2,2})^2 T L^{(T)}(\bar{\gamma}_2)^2$$
$$\times d\alpha_{1,2}\cdots d\alpha_{2,2}d\lambda_1 d\lambda_2$$

and further by integrating over λ_1, λ_2 and the variables in P_1

$$\leq \frac{C}{T^3}\int_{\Pi} L^{(T)}(\bar{\gamma}_2)^2 d\bar{\gamma}_2 \leq \frac{C}{T^2}.$$

Otherwise, one of the sets is of the form $\{\alpha,-\alpha\}$ and hence corresponds to a zero term in (19). The same applies to partitions with three or four non-hooked subpartitions.

Lemma 7 *Suppose that Assumptions 1–4 hold. Furthermore, let $M_T\to\infty$ and $M_T^2/T\to 0$. Then, if $(X_{t,1})$ and $(X_{t,2})$ are uncorrelated, the cumulants of kth order of $\tilde{S}_T^{(2)}$ satisfy*

$$cum_k\{\tilde{S}_T^{(2)}\}=o\left(\frac{M_T^{k/2}}{T^k}\right)$$

for all $k\geq 3$.

Proof Let $\Delta(\lambda)$ be defined as in the proof of Lemma 4. Then, by (12), we obtain for the kth order cumulant similarly as for the variance in the previous lemma

$$\left|\mathrm{cum}_k\{\tilde{S}_T^{(2)}\}\right| \leq \frac{C}{T^{2k}} \sum_{i.p.} \Big| \int_{\Pi^{3k}} \prod_{j=1}^{k} \Big[g(\lambda_j) w^{(T)}(\lambda_j - \alpha_{j,1}) w^{(T)}(\lambda_j - \alpha_{j,2}) \Big] \times \prod_{j=1}^{m} H_{p_j,T}(\bar{\gamma}_j) f_{\nu_{j,1}\ldots\nu_{j,p_j}}(\gamma_{j,1}, \ldots, \gamma_{j,p_j-1}) \times d\alpha_{1,1} \cdots d\alpha_{k,2} d\lambda_1 \cdots d\lambda_k \Big| + R_T, \tag{20}$$

where the remainder term R_T is of smaller order than the main term. Therefore it suffices to show that each summand of the sum in the main term asymptotically tends to zero with rate $M_T^{k/2}/T^k$.

Let $\{P_1, \ldots, P_m\}$ be an indecomposable partition of table (10) which consists of s non-hooked subpartitions. Since any subpartition that covers only one variable corresponds to a zero factor $f_{12}(\alpha_{i,j})$ in the above product, we can assume that every subpartition covers at least two variables. Then the corresponding summand in (20) is bounded by

$$\frac{C}{T^{2k} M_T^{2k}} \int_{\Pi^{3k}} \prod_{j=1}^{k} \Big[L^{(M_T)}(\lambda_j - \alpha_{j,1})^2 \, L^{(M_T)}(\lambda_j - \alpha_{j,2})^2 \Big] \times \prod_{j=1}^{m-u} L^{(T)}(\bar{\gamma}_j) d\alpha_{1,1} \cdots d\alpha_{k,1} d\alpha_{1,2} \cdots d\alpha_{k-u,2} d\lambda_1 \cdots d\lambda_k.$$

Next, suppose there are s' subpartitions $\{P_{i_1}, \ldots, P_{i_r}\}$ such that $p_{i_j} = 2$ for all $j = 1, \ldots, r$. Then, according to Lemma 3, we can select two variables $\alpha_{l_{j,1}, i_{j,1}}$ and $\alpha_{l_{j,2}, i_{j,2}}$ such that the set of row indices $\{(l_{1,1}, l_{1,2}), \ldots, (l_{s',1}, l_{s',2})\}$ does not contain more than one circle.

Similarly for each of the remaining $s - s'$ subpartitions, we can choose two variables, such that the remaining variables in the set satisfy the conditions of Lemma 2. Now we can bound $L^{(M_T)}(\cdot)^2$ by M_T^2 for each of the $2k - 2s$ variables that are left. Thus, integrating over these variables, we have by Lemma 2

$$C \log(T)^{\rho} \frac{M_T^{2k-4s}}{T^{2k-s+s'}} \int_{\Pi^{k+2s}} \prod_{j=1}^{s} \Big[L^{(M_T)}(\lambda_{l_{j,1}} - \alpha_{l_{j,1}, i_{j,1}})^2 L^{(M_T)}(\lambda_{l_{j,2}} - \alpha_{l_{j,2}, i_{j,2}})^2 \Big] \times \prod_{j=1}^{s'} L^{(T)}(\alpha_{l_{j,1}, i_{j,1}} \pm \alpha_{l_{j,2}, i_{j,2}})^2 d\alpha_{l_{1,1}, i_{1,1}} \cdots d\alpha_{l_{s,2}, i_{s,2}} d\lambda_1 \cdots d\lambda_k$$

for some $\rho \in \mathbb{N}$. Integrating over $\alpha_{l_{s'+1,1}, i_{s'+1,1}}, \ldots, \alpha_{l_{s,2}, i_{s,2}}$, we further obtain

$$C \log(T)^{\rho} \frac{M_T^{2(k-s-s')}}{T^{2k-s+s'}} \int_{\Pi^{k+2s'}} \prod_{j=1}^{s'} \Big[L^{(M_T)}(\lambda_{l_{j,1}} - \alpha_{l_{j,1}, i_{j,1}})^2 L^{(M_T)}(\lambda_{l_{j,2}} - \alpha_{l_{j,2}, i_{j,2}})^2 \Big]$$

$$\times \prod_{j=1}^{s'} L^{(T)}(\alpha_{l_{j,1},i_{j,1}} \pm \alpha_{l_{j,2},i_{j,2}})^2 d\alpha_{l_{1,1},i_{1,1}} \cdots d\alpha_{l_{s',2},i_{s',2}} d\lambda_1 \cdots d\lambda_k.$$

Integrating over the remaining $\alpha_{l_{j,1},i_{j,1}}$, we get

$$C \log(T)^\rho \frac{M_T^{2(k-s)-s'}}{T^{2k-s}} \int_{\Pi^k} \prod_{j=1}^{s'} L^{(M_T)}(\lambda_{l_{j,2}} \mp \lambda_{l_{j,1}})^2 d\lambda_1 \cdots d\lambda_k.$$

Since the pairs $(l_{j,1}, l_{j,2})$ where chosen such that at most one circle exists, integration over $\lambda_1, \ldots, \lambda_k$ yields an additional factor of order $O(M_T^{s'+1})$. Thus, for any partition $\{P_1, \ldots, P_M\}$ with s non-hooked subpartitions, the corresponding summand in (20) is bounded by

$$C \frac{M_T^{k/2}}{T^k} \left(\frac{M_T^2}{T}\right)^{k-s+1/2-k/4} \frac{\log(T)^\rho}{T^{k/2-1}}.$$

Since $s \leq k$ and $k \geq 3$, this is of order $o\left(M_T^{k/2}/T^k\right)$.

Proof of Theorem 6 Under the local alternative with $f_{12}^{(T)}(\lambda) = c_T\, g(\lambda)$, the statistic $\tilde{S}_T$ can be approximated by

$$\tilde{S}_T^{(2)} = c_T\, \nu(g) + \int_\Pi \Delta(\lambda) \left(\tilde{f}_{12}^{(T)}(\lambda)\, \tilde{f}_{21}^{(T)}(\lambda) - f_{12}^{(T)}(\lambda)\, f_{21}^{(T)}(\lambda)\right) d\lambda.$$

with $\tilde{S}_T - \tilde{S}_T^{(2)} = o_P(c_T) + O_P\left((M_T/T)^{3/2}\right) = o_P(M_T^{1/2}/T)$. When we evaluate the mean of $\tilde{S}_T^{(2)}$ under the local alternative, the second term in (18) cancels with the constant term in $\tilde{S}_T^{(2)}$ while all other steps remain the same.

For the convergence of the cumulants of second and higher order, we note that

$$\begin{aligned}&\int_\Pi w^{(T)}(\lambda - \alpha)\, \mathrm{cum}\{d_1^{(T)}(\alpha), d_2^{(T)}(-\alpha)\}\\ &\quad = 2\pi H_{2,T} \int_\Pi w^{(T)}(\lambda - \alpha)\, f_{12}(\alpha) d\lambda + O(1) = o(c_T\, T).\end{aligned}$$

For $c_T = M_T^{1/4}/T^{1/2}$, it can be shown that all terms corresponding to partitions with subpartitions covering only one variable are of order $o(M_T^{k/2}/T^k)$. The remaining steps are the same with slight modifications.

If we consider S_T instead of $\tilde{S}_T$, we have $S_T = S_T^{(2)} + o_P(M_T^{1/2}/T)$ with

$$S_T^{(2)} = \frac{2\pi}{T} \sum_{j=0}^{T-1} \frac{\hat{f}_{12}(\lambda_j)\, \hat{f}_{21}(\lambda_j)}{2\pi\, f_{11}(\lambda_j)\, f_{22}(\lambda_j)}$$

and $\lambda_j = 2\pi j/T$. The proofs of Lemmas 4, 6, and 7 remain the same with integrals replaced by sums. For the application of the $L^{(T)}$ function, note that it can be shown that

$$\sum_{j=0}^{T-1} L^{(T)}(\lambda_j)^r \leq C \int_{\Pi} L^{(T)}(\lambda)^r \, d\lambda$$

for some constant C independent of T. Thus the upper bounds remain unchanged in the discrete case. Finally, since w, f_{ab}, and $\Delta = (2\pi f_{11} f_{22})^{-1}$ are all Lipschitz continuous, all sums in the final expressions converge to integrals with a rate of $o(T^{-1})$. This concludes the proof of Theorems 5 and 6.

Proof of Theorem 7 The result is a straightforward modification of Theorems 5 and 6.

References

Beltrão KI, Bloomfield P (1987) Determining the bandwidth of a kernel spectrum estimate. J Time Ser Anal 8:21–38

Brillinger DR (1981) Time series: data analysis and theory. McGraw Hill, New York

Bühlmann P (1996) Locally adaptive lag-window spectral estimation. J Time Ser Anal 17:247–270

Dahlhaus R (1983) Spectral analysis with tapered data. J Time Ser Anal 4:163–175

Dahlhaus R (1990) Nonparametric high resolution spectral estimation. Probab Theory Relat Fields 85:147–180

Dahlhaus R, Eichler M (2003) Causality and graphical models in time series analysis. In: Green P, Hjort N, Richardson S (eds) Highly structured stochastic systems. University Press, Oxford

Dahlhaus R, Eichler M, Sandkühler J (1997) Identification of synaptic connections in neural ensembles by graphical models. J Neurosci Methods 77:93–107

Duchesne P, Roy R (2003) Robust test for independence of two time series. Stat Sinica 13:827–852

El Himdi K, Saidi A (1997) Tests for noncorrelation of two multivariate ARMA time series. Can J Stat 25:233–256

Fortin I, Kuzmics C (2000) Optimal window width choice in spectral density estimation. J Stat Comput Simul 67:109–131

Hallin M, Saidi A (2005) Testing non-correlation and non-causality between multivariate ARMA time series. J Time Ser Anal 26:83–105

Haugh LD (1976) Checking the independence of two covariance-stationary time series: a univariate residual covariance approach. J Am Stat Assoc 71:378–385

Hong Y (1996) Testing for independence between two covariance stationary time series. Biometrika 83:615–625

Koch PD, Yang SS (1986) A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. J Am Stat Assoc 81:533–544

Lee TCM 2001 A stabilized bandwidth selection method for kernel smoothing of the periodogram. Signal Process 81:419–430

Priestley MB (1981) Spectral analysis and time series. vol 1, Academic, London

Taniguchi M, Puri ML, Kondo M (1996) Nonparametric approach for non-gaussian vector stationary processes. J Multivar Anal 56:259–283

Timmer J, Lauk M, Köster B, Hellwig B, Häußler S, Guschlbauer B, Radt V, Eichler M, Deuschl G, Lücking CH 2000 Cross-spectral analysis of tremor time series. Int J Bifurcat Chaos 10:2595–2610