Bernoulli **19**(5B), 2013, 2715–2749
DOI: 10.3150/12-BEJ472

A test for stationarity based on empirical processes

PHILIP PREUß*, MATHIAS VETTER** and HOLGER DETTE†

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: *philip.preuss@ruhr-uni-bochum.de; **mathias.vetter@ruhr-uni-bochum.de;
†holger.dette@ruhr-uni-bochum.de

In this paper we investigate the problem of testing the assumption of stationarity in locally stationary processes. The test is based on an estimate of a Kolmogorov–Smirnov type distance between the true time varying spectral density and its best approximation through a stationary spectral density. Convergence of a time varying empirical spectral process indexed by a class of certain functions is proved, and furthermore the consistency of a bootstrap procedure is shown which is used to approximate the limiting distribution of the test statistic. Compared to other methods proposed in the literature for the problem of testing for stationarity the new approach has at least two advantages: On one hand, the test can detect local alternatives converging to the null hypothesis at any rate $g_T \to 0$ such that $g_T T^{1/2} \to \infty$, where T denotes the sample size. On the other hand, the estimator is based on only one regularization parameter while most alternative procedures require two. Finite sample properties of the method are investigated by means of a simulation study, and a comparison with several other tests is provided which have been proposed in the literature.

Keywords: bootstrap; empirical spectral measure; goodness-of-fit tests; integrated periodogram; locally stationary process; non-stationary processes; spectral density

1. Introduction

Most literature in time series analysis assumes that the underlying process is second-order stationary. This assumption allows for an elegant development of powerful statistical methodology like parameter estimation or forecasting techniques, but is often not justified in practice. In reality, most processes change their second-order characteristics over time and numerous models have been proposed to address this feature. Out of the large literature, we mention exemplarily the early work on this subject by Priestley [25], who considered oscillating processes. More recently, the concept of locally stationary processes has found considerable attention, because in contrast to other proposals it allows for a meaningful asymptotic theory, which is essential for statistical inference in such models. The class of locally stationary processes was introduced by Dahlhaus [7] and particular important examples are time varying ARMA models.

While many estimation techniques for locally stationary processes were developed (see Neumann and von Sachs [20], Dahlhaus, Neumann and von Sachs [10], Chiann and Morettin [5], Dahlhaus and Polonik [11], Dahlhaus and Subba Rao [13], Van Bellegem and von Sachs [27] or Palma and Olea [21] among others), goodness-of-fit testing has found much less attention although its importance was pointed out by many authors. von Sachs and Neumann [29] proposed a method to test the assumption of stationarity, which is based on the estimation of wavelet coefficients by a localised version of the periodogram. Paparoditis [22] and Paparoditis [23] used an L_2

distance between the true spectral density and its best approximation through a stationary spectral density to measure deviations from stationarity, and most recently Dwivedi and Subba Rao [15] developed a Portmanteau type test statistic to detect non-stationarity. However, besides the choice of a window width for the localised periodogram which is inherent in essentially any statistical inference for locally stationary processes, all these concepts require the choice of at least one additional regularization parameter. For example, the procedure proposed in Sergides and Paparoditis [26] relies on an additional smoothing bandwidth for the estimation of the local spectral density. It was pointed out therein that it is the choice of this particular tuning parameter that influences the results of the statistical analysis substantially.

Recently, Dette, Preuss and Vetter [14] proposed a test for stationarity which is based on an L_2 distance between the true spectral density and its best stationary approximation and which does not require the choice of that additional regularization parameter. Roughly speaking, these authors proposed to estimate the L_2 distance considered by Paparoditis [22] by calculating integrals of powers of the spectral density directly via Riemann sums of the periodogram. With this idea, Dette, Preuss and Vetter [14] avoided the integration of the smoothed periodogram, as it was done in Paparoditis [22] or Paparoditis [23]. In a comprehensive simulation study it was shown that this method is superior compared to the other tests, no matter how the additional smoothing bandwidths in these procedures are chosen.

Although the test proposed by Dette, Preuss and Vetter [14] has attractive features, it can only detect local alternatives converging to the null hypothesis at a rate $T^{-1/4}$, where T here and throughout the paper denotes the sample size. It is the aim of the present paper to develop a test for stationarity in locally stationary processes which is at first able to detect alternatives converging to the null hypothesis at the rate $g_T \to 0$ such that $g_T T^{1/2} \to \infty$ and is secondly based on the concept in Dette, Preuss and Vetter [14] for which no additional smoothing bandwidth is needed. For this purpose, we employ a Kolmogorov–Smirnov type test statistic to estimate a measure of deviation from stationarity, which is defined by

$$D := \sup_{(v,\omega)\in[0,1]^2} \left|D(v,\omega)\right|,$$

where for all $(v,\omega) \in [0,1]^2$ we set

$$D(v,\omega) := \frac{1}{2\pi}\left(\int_0^v \int_0^{\pi\omega} f(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u - v\int_0^{\pi\omega}\int_0^1 f(u,\lambda)\,\mathrm{d}u\,\mathrm{d}\lambda\right) \tag{1.1}$$

and where $f(u,\lambda)$ denotes the time varying spectral density. Note that the quantity D is identically zero if the process is stationary, that is, if $f(u,\lambda)$ is does not depend on u. The consideration of functionals of the form (1.1) for the construction of a test for stationarity is natural and was already suggested by Dahlhaus [9]. In particular, Dahlhaus and Polonik [12] proposed an estimator of this quantity which is based on the integrated (with respect to the Lebesgue measure) pre-periodogram. However, in applications Riemann sums are used to approximate the integral and therefore the approach proposed by these authors is not directly implementable. In particular, it is pointed out in Example 2.7 of Dahlhaus [9] that the asymptotic properties of an estimator based on Riemann approximation have been an open problem so far. See the discussion at the end of Section 2 for more details.

In Section 2, we introduce an alternative stochastic process, say $\{\hat{D}_T(v,w)\}_{(v,w)\in[0,1]^2}$, which is based on a summation of the localised periodogram and serves as an estimate of $\{D(v,w)\}_{(v,w)\in[0,1]^2}$. The proposed statistic does neither require integration of the localised periodogram with respect to an absolutely continuous measure nor the problematic choice of a second regularization parameter. Weak convergence of a properly standardized version of $\hat{D}_T$ to a Gaussian process is established under the null hypothesis, local and fixed alternatives, giving a consistent estimate of D. The distribution of the limiting process depends on certain features of the data generating process which are difficult to estimate. Therefore, the second purpose of this paper is the development of an AR(∞) bootstrap method and a proof of its consistency. See Section 3 for details. We also provide a solution of the problem mentioned in the previous paragraph and prove weak convergence of a Riemann approximation for the integrated pre-periodogram proposed by Dahlhaus [9], which is Theorem 2.2 in the following section. As a result, we obtain two empirical processes estimating the function D defined in (1.1) which differ by the use of localised periodogram and pre-periodogram in the Riemann approximations. In Section 4, we investigate their finite sample properties by means of a simulation study. Although the estimator based on the pre-periodogram does not require the specification of any regularization parameter at all, it is demonstrated that it yields substantially less power compared to the statistic based on the localised periodogram. Additionally, it is shown that the latter method is extremely robust with respect to different choices of the window width which is used for the calculation of the localised periodogram. Moreover, we also provide a comparison with the tests proposed in Paparoditis [23], Dwivedi and Subba Rao [15] and Dette, Preuss and Vetter [14] and show that the new proposal performs better in many situations. Finally, we present a data example, and for the sake of a transparent presentation of the results all technical details are deferred to the Appendix.

2. The test statistic

Following Dahlhaus and Polonik [12], we define a locally stationary process via a sequence of stochastic processes $\{X_{t,T}\}_{t=1,\dots,T}$ which exhibit a time varying MA(∞) representation, namely

$$X_{t,T} = \sum_{l=-\infty}^{\infty} \psi_{t,T,l} Z_{t-l}, \qquad t = 1, \dots, T, \tag{2.1}$$

where the random variables Z_t are independent identically standard normal distributed random variables. Since the coefficients $\psi_{t,T,l}$ are in general time dependent, each process $\{X_{t,T}\}_{t=1,\dots,T}$ is typically not stationary. To ensure that the process shows approximately stationary behavior on a small time interval, we impose that there exist twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$, $l \in \mathbb{Z}$, such that

$$\sum_{l=-\infty}^{\infty} \sup_{t=1,\dots,T} \left|\psi_{t,T,l} - \psi_l(t/T)\right| = \mathrm{O}(1/T) \tag{2.2}$$

as $T \to \infty$. Furthermore, we assume that the following technical conditions

$$\sum_{l=-\infty}^{\infty} \sup_{u\in[0,1]} \left|\psi_l(u)\right||l| < \infty, \tag{2.3}$$

$$\sum_{l=-\infty}^{\infty} \sup_{u\in[0,1]} \left|\psi_l'(u)\right| < \infty, \tag{2.4}$$

$$\sum_{l=-\infty}^{\infty} \sup_{u\in[0,1]} \left|\psi_l''(u)\right| < \infty \tag{2.5}$$

are satisfied, which are in general rather mild. See Dette, Preuss and Vetter [14] for a discussion. Note that variables Z_t with time varying variance $\sigma^2(t/T)$ can be included in the model by choosing the coefficients $\psi_{t,T,l}$ in (2.1) appropriately.

Set

$$\psi\bigl(u, \exp(-\mathrm{i}\lambda)\bigr) := \sum_{l=-\infty}^{\infty} \psi_l(u) \exp(-\mathrm{i}\lambda l).$$

Then the function

$$f(u, \lambda) = \frac{1}{2\pi}\left|\psi\bigl(u, \exp(-\mathrm{i}\lambda)\bigr)\right|^2$$

is well defined and called the time varying spectral density of $\{X_{t,T}\}_{t=1,\ldots,T}$, see Dahlhaus [7]. It is continuous by assumption and can roughly be estimated by a local periodogram. To be precise, we assume without loss of generality that the total sample size T can be decomposed as $T = NM$, where N and M are integers and N is even. Furthermore, we define

$$I_N^X(u, \lambda) := \frac{1}{2\pi N}\left|\sum_{s=0}^{N-1} X_{\lfloor uT\rfloor - N/2+1+s,T} \exp(-\mathrm{i}\lambda s)\right|^2,$$

which is the local periodogram at time u proposed by Dahlhaus [8]. Here, we have set $X_{j,T} = 0$, if $j \notin \{1, \ldots, T\}$. This is the usual periodogram computed from the observations $X_{\lfloor uT\rfloor - N/2+1,T}, \ldots, X_{\lfloor uT\rfloor + N/2,T}$. The arguments employed in the Appendix show that

$$\mathbb{E}\bigl(I_N^X(u, \lambda)\bigr) = f(u, \lambda) + \mathrm{O}(1/N) + \mathrm{O}(N/T),$$

and therefore the statistic $I_N^X(u, \lambda)$ is an asymptotically unbiased estimator for the spectral density if $N \to \infty$ and $N = \mathrm{o}(T)$. However, $I_N^X(u, \lambda)$ is not consistent just as the usual periodogram.

We now consider an empirical version of the function $D(v, \omega)$ defined in (1.1), that is,

$$\hat{D}_T(v, \omega) := \frac{1}{T}\sum_{j=1}^{\lfloor vM\rfloor}\sum_{k=1}^{\lfloor \omega N/2\rfloor} I_N^X(u_j, \lambda_k) - \frac{\lfloor vM\rfloor}{M}\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \omega N/2\rfloor} I_N^X(u_j, \lambda_k), \tag{2.6}$$

where the points

$$u_j := \frac{t_j}{T} := \frac{N(j-1)+N/2}{T}, \qquad j = 1, \dots, M,$$

define an equidistant grid of the interval [0, 1] and

$$\lambda_k := \frac{2\pi k}{N}, \qquad k = 1, \dots, \frac{N}{2},$$

denote the Fourier frequencies. It follows from the proof of Theorem 2.1 in the Appendix that for every $v \in [0, 1]$ and $\omega \in [0, 1]$ we have

$$\begin{aligned}\mathbb{E}\big(\hat{D}_T(v,\omega)\big) &= \frac{1}{T}\sum_{j=1}^{\lfloor vM \rfloor}\sum_{k=1}^{\lfloor \omega N/2 \rfloor} f(u_j, \lambda_k) \\ &\quad - \frac{\lfloor vM \rfloor}{M}\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \omega N/2 \rfloor} f(u_j, \lambda_k) + \mathrm{O}(1/N) + \mathrm{O}\big(N^2/T^2\big) \\ &= D(v,\omega) + \mathrm{O}(1/N) + \mathrm{O}(N/T),\end{aligned}$$

where the latter identity is due to the approximation error of the Riemann sum. This error can be improved, if we replace $D(v, \omega)$ by its discrete time approximation, that is,

$$D_{N,M}(v,\omega) := D\left(\frac{\lfloor vM \rfloor}{M}, \frac{\lfloor \omega N/2 \rfloor}{N/2}\right)$$

for which the representation

$$\mathbb{E}\big(\hat{D}_T(v,\omega)\big) = D_{N,M}(v,\omega) + \mathrm{O}(1/N) + \mathrm{O}\big(N^2/T^2\big) \tag{2.7}$$

holds. The approximation error of the Riemann sum in (2.7) becomes smaller due to the choice of the midpoints u_j. The rate of convergence will be $T^{-1/2}$ later on, so we need the $\mathrm{O}(\cdot)$-terms to vanish asymptotically after multiplication with $\sqrt{T}$. Therefore, we define an empirical spectral process by

$$\hat{G}_T(v,\omega) := \sqrt{T}\left(\frac{1}{T}\sum_{j=1}^{\lfloor vM \rfloor}\sum_{k=1}^{\lfloor \omega N/2 \rfloor} I_N^X(u_j,\lambda_k) - \frac{\lfloor vM \rfloor}{M}\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \omega N/2 \rfloor} I_N^X(u_j,\lambda_k) - D_{N,M}(v,\omega)\right)$$

and assume

$$N \to \infty, \qquad M \to \infty, \qquad \frac{T^{1/2}}{N} \to 0, \qquad \frac{N}{T^{3/4}} \to 0. \tag{2.8}$$

Our first result specifies the asymptotic properties of the empirical process $(\hat{G}_T(v,\omega))_{(v,\omega)\in[0,1]^2}$, both under the null hypothesis and under a fixed alternative. The null hypothesis of stationarity is formulated as

$$H_0 : f(u,\lambda) \text{ is independent of } u, \tag{2.9}$$

which is a little different from genuine second-order stationarity, since it only means that the coefficients $\psi_{t,T,l}$ in (2.1) can be approximated by time independent terms ψ_l. Thanks to the continuity of the time varying spectral density, the alternative corresponds to the property that there is some λ such that $u \mapsto f(u,\lambda)$ is not a constant function. Finally, the symbol $\Rightarrow$ denotes weak convergence in $[0,1]^2$.

Theorem 2.1. *Suppose we have a locally stationary process as defined in* (2.1) *with independent and standard normal innovations* Z_t. *Furthermore assume that the assumptions* (2.2)–(2.5) *and* (2.8) *are satisfied. Then as* $T \to \infty$ *we have*

$$\big(\hat{G}_T(v,\omega)\big)_{(v,\omega)\in[0,1]^2} \Rightarrow \big(G(v,\omega)\big)_{(v,\omega)\in[0,1]^2}, \tag{2.10}$$

where $(G(v,\omega))_{(v,\omega)\in[0,1]^2}$ *is a Gaussian process with mean zero and covariance structure*

$$\begin{aligned}&\operatorname{Cov}\big(G(v_1,\omega_1),G(v_2,\omega_2)\big)\\&\quad=\frac{1}{2\pi}\int_0^1\int_0^{\pi\min(\omega_1,\omega_2)}\big(1_{[0,v_1]}(u)-v_1\big)\big(1_{[0,v_2]}(u)-v_2\big)f^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u.\end{aligned}$$

Under the null hypothesis, we have $D_{N,M}(v,\omega)=0$ for all $N,M\in\mathbb{N}$ and for all $v,\omega\in[0,1]$. Therefore, we obtain

$$\big(\sqrt{T}\hat{D}_T(v,\omega)\big)_{(v,\omega)\in[0,1]^2} \Rightarrow \big(G(v,\omega)\big)_{(v,\omega)\in[0,1]^2},$$

which yields

$$\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}\big|\hat{D}_T(v,\omega)\big| \xrightarrow{D} \sup_{(v,\omega)\in[0,1]^2}\big|G(v,\omega)\big| \tag{2.11}$$

under the null hypothesis (2.9). An asymptotic level α test is then obtained by rejecting the null hypothesis of stationarity whenever $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ exceeds the $(1-\alpha)\%$ quantile of the distribution of the random variable $\sup_{(v,\omega)\in[0,1]^2}|G(v,\omega)|$. On the other hand, under the alternative there is a pair (v,ω) such that $D(v,\omega)\neq 0$. The fact that $D_{N,M}$ converges uniformly to D together with Theorem 2.1 yields consistency of this test. Note also that under the null hypothesis H_0 the covariance structure of the limiting process in Theorem 2.1 simplifies to

$$\operatorname{Cov}\big(G(v_1,\omega_1),G(v_2,\omega_2)\big)=\frac{\min(v_1,v_2)-v_1v_2}{2\pi}\int_0^{\pi\min(\omega_1,\omega_2)} f^2(\lambda)\,\mathrm{d}\lambda \tag{2.12}$$

and depends on the unknown spectral density f. In order to avoid the estimation of the integral over the squared spectral density, we propose to approximate the quantiles of the limiting distribution by an AR(∞) bootstrap, which will be described in the following section.

An alternative estimator for the time varying spectral density is given by

$$J_T(u,\lambda) := \frac{1}{2\pi} \sum_{k\colon 1\le \lfloor uT+1/2\pm k/2\rfloor \le T} X_{\lfloor uT+1/2+k/2\rfloor} X_{\lfloor uT+1/2-k/2\rfloor} \exp(-\mathrm{i}\lambda k),$$

which is called the pre-periodogram (see Neumann and von Sachs [20]). As for the usual periodogram, it is asymptotically unbiased, but again not consistent. Based on this statistic, we define an alternative process by

$$\begin{aligned}\hat{H}_T^1(v,\omega) := \sqrt{T}\Biggl(\frac{1}{T^2}\sum_{j=1}^{\lfloor vT\rfloor}\sum_{k=1}^{\lfloor \omega T/2\rfloor} J_T(j/T,\lambda_{k,T}) \\ - \frac{\lfloor vT\rfloor}{T^3}\sum_{j=1}^{T}\sum_{k=1}^{\lfloor \omega T/2\rfloor} J_T(j/T,\lambda_{k,T}) - D(v,\omega)\Biggr),\end{aligned} \tag{2.13}$$

where the Fourier frequencies become $\lambda_{k,T} = 2\pi k/T$ now. Convergence of the finite dimensional distributions of the process $(H_T^1(v,\omega))_{(v,\omega)\in[0,1]^2}$ to the ones of the limiting process $(G(v,\omega))_{(v,\omega)\in[0,1]^2}$ has already been shown in Dahlhaus [9]. Tightness can be shown using similar arguments as given in the Appendix for the proof of Theorem 2.1, which are not stated here for the sake of brevity. As a consequence, we obtain the following result.

Theorem 2.2. *If the assumptions of Theorem* 2.1 *are satisfied, then as* $T\to\infty$ *we have*

$$\bigl(\hat{H}_T^1(v,\omega)\bigr)_{(v,\omega)\in[0,1]^2} \Rightarrow \bigl(G(v,\omega)\bigr)_{(v,\omega)\in[0,1]^2},$$

where $(G(v,\omega))_{(v,\omega)\in[0,1]^2}$ *is the Gaussian process defined in Theorem* 2.1.

Because the use of $\hat{H}_T^1(v,\omega)$ instead of $\hat{G}_T(v,\omega)$ does not require the choice of the quantity N, which specifies the number of observations used for the calculation of the local periodogram, it might be appealing to construct a Kolmogorov–Smirnov type test for stationarity on the basis of this process. However, we will demonstrate in Section 4 by means of a simulation study that for realistic sample sizes the method which employs the pre-periodogram is clearly outperformed by the approach based on the local periodogram. Our numerical results also show that the use of the local periodogram is not very sensitive with respect to the choice of the regularization parameter N either, and therefore we strictly recommend to use the latter approach when constructing a Kolmogorov–Smirnov test.

Remark 2.3. The convergence of a modified version of the process (2.13) to the limiting Gaussian process $(G(v,\omega))_{(v,\omega)\in[0,1]^2}$ of Theorem 2.1 was shown in Dahlhaus and Polonik [12], where the Riemann sum over the Fourier frequencies was replaced by the integral with respect to the Lebesgue measure. More precisely, these authors considered the process

$$\big(\hat{H}_T^2(v,\omega)\big)_{(v,\omega)\in[0,1]^2} := \frac{1}{2\pi\sqrt{T}}\Bigg(\sum_{j=1}^{\lfloor vT\rfloor}\int_0^{\pi\omega} J_T(j/T,\lambda)\,\mathrm{d}\lambda - v\sum_{j=1}^{T}\int_0^{\pi\omega} J_T(j/T,\lambda)\,\mathrm{d}\lambda - D(v,\omega)\Bigg)_{(v,\omega)\in[0,1]^2}$$

instead of $(H_T^1(v,\omega))_{(v,\omega)\in[0,1]^2}$ and proved its weak convergence. This is a rather typical result, as many other asymptotic results are only shown for the integral (instead of the sum over the Fourier coefficients) over the local periodogram or the pre-periodogram; see, for example, Dahlhaus [8] or Paparoditis [23]. The transition from these results to analogue statements for the corresponding Riemann approximations is by no means obvious. For example, although it is appealing to assume that

$$\int_0^{\pi} I_N^X(u,\lambda)\,\mathrm{d}\lambda = \frac{2\pi}{N}\sum_{k=1}^{N/2} I_N^X(u,\lambda_k) + \mathrm{O}(1/N)$$

holds because of the Riemann approximation error, this identity is in general not valid, as the derivative $\partial I_N^X(u,\lambda)/\partial\lambda$ is not uniformly bounded in N. A demonstrative explanation of this fact is that $I_N^X(u,\lambda_{k_1})$ and $I_N^X(u,\lambda_{k_2})$ are asymptotically independent whenever $k_1 \neq k_2$. Thus in general asymptotic results for integrated local periodogram or pre-periodogram cannot be directly transferred to the corresponding Riemann approximations. These difficulties were also explicitly pointed out in Example 2.7 of Dahlhaus [9]. Note further that asymptotic tightness has neither been studied for an integrated nor for a summarized local periodogram in the literature so far.

Remark 2.4. Suppose that we are in the situation of local alternatives, that is, we have

$$f_T(u,\lambda) = f(\lambda) + g_T k(u,\lambda) \tag{2.14}$$

for some deterministic sequence g_T and an appropriate function k such that (2.14) defines a time varying spectral density. Note that a locally stationary process with this specific spectral density can easily be constructed through the equation

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(\mathrm{i}\lambda t) A_T(t/T,\lambda)\,\mathrm{d}\xi(\lambda),$$

where ξ is an orthogonal increment Gaussian process and $A_T(u,\lambda)$ is a function such that $f_T(u,\lambda) = |A_T(u,\lambda)|^2$. See Dahlhaus [8].

A careful inspection of the proofs in the Appendix shows that (2.10) with centering term $D_{N,M}(v,\omega) = 0$ and asymptotic covariance (2.12) also holds in the case where $g_T = \mathrm{o}(1/\sqrt{T})$.

Moreover, if $g_T = 1/\sqrt{T}$ an analogue of Theorem 2.1 can be obtained where the centering term $D_{N,M}(v,\omega)$ in the definition of $\hat{G}_T(v,\omega)$ is replaced by

$$D_{N,M,k}(v,\omega) = \frac{1}{2\pi\sqrt{T}}\biggl(\int_0^{\lfloor vM\rfloor/M}\int_0^{2\pi\lfloor\omega N/2\rfloor/N} k(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u \\ - \frac{\lfloor vM\rfloor}{M}\int_0^{2\pi\lfloor\omega N/2\rfloor/N}\int_0^1 k(u,\lambda)\,\mathrm{d}u\,\mathrm{d}\lambda\biggr),$$

which is the original $D_{N,M}$ but with $T^{-1/2}k(u,\lambda)$ playing the role of $f(u,\lambda)$. In this case, the appropriately centered process converges weakly to a Gaussian process $\{G(v,\omega)\}_{(v,\omega)\in[0,1]^2}$ with covariance structure given by (2.12) as well. A similar comment applies to the process $\hat{H}_T^1$ defined in (2.13). This means that the tests based on the processes $\hat{G}_T$ and $\hat{H}_T^1$ can detect alternatives converging to the null hypothesis at any rate $g_T \to 0$ such that $g_T T^{1/2} \to \infty$. In contrast, the proposal of Dette, Preuss and Vetter [14] is based on an L_2 distance between $f(u,\lambda)$ and $\int_0^1 f(v,\lambda)\,\mathrm{d}v$ and is therefore only able to detect alternatives converging to the null hypothesis at a rate $T^{-1/4}$.

Remark 2.5. In Theorems 2.1 and 2.2, we assume the existence of second order derivatives for the approximating functions $\psi_l(u)$. Nevertheless, it is straightforward to show that our test also detects fixed alternatives in which the $\psi_l(u)$ admit a finite number of points of discontinuity. We furthermore conjecture that the constraints in Theorems 2.1 and 2.2 can be weakened to some kind of condition on the total variation of $\psi_l(u)$ as in Definition 2.1 in Dahlhaus and Polonik [12].

3. Bootstrapping the test statistic

To approximate the limiting distribution of $\sup_{(v,\omega)\in[0,1]^2}|G(v,\omega)|$, we will employ an AR($\infty$) bootstrap approximation, which was introduced by Kreiß [17]. To ensure consistency of the bootstrap procedure described later, we have to consider the stationary process Y_t with spectral density $\lambda \mapsto \int_0^1 f(u,\lambda)\,\mathrm{d}u$ first, which coincides with $X_{t,T}$ in case the latter process is stationary. We have to impose the following main assumption.

Assumption 3.1. *We assume that the spectral density $\lambda \mapsto \int_0^1 f(u,\lambda)\,\mathrm{d}u$ is strictly positive and that the process Y_t has an* AR(∞) *representation, that is,*

$$Y_t = \sum_{j=1}^{\infty} a_j Y_{t-j} + Z_t^{\mathrm{AR}}, \tag{3.1}$$

where $(Z_j^{\mathrm{AR}})_{j\in\mathbb{Z}}$ denotes a Gaussian white noise process with some variance $\sigma^2 > 0$ and the sequence $(a_j)_{j\in\mathbb{N}}$ of coefficients satisfies $\sum_{j=1}^{\infty}|a_j| < \infty$ and

$$1 - \sum_{j=1}^{\infty} a_j z^j \neq 0 \qquad \textit{for } |z| \le 1. \tag{3.2}$$

Note that $(Y_t)_{t\in\mathbb{Z}}$ possesses an MA(∞) representation

$$Y_t = \sum_{l=-\infty}^{\infty} \psi_l Z_{t-l}, \tag{3.3}$$

where the Z_t are the same as in (2.1) and the ψ_l are some appropriately defined constants. The random variables Z_t^{AR} in (3.1) do not necessarily coincide with the Z_t from (3.3), even though this could be ensured by assuming that the MA(∞) representation in (3.3) corresponds to the Wold representation of Y_t. See, for example, Kreiss, Paparoditis and Politis [19] for a comprehensive illustration.

We have to introduce a second class of stationary processes, namely $(Y_t^{\mathrm{AR}}(p))_{t\in\mathbb{Z}}$ for arbitrary integer p, which is the process defined through

$$Y_t^{\mathrm{AR}}(p) = \sum_{j=1}^{p} a_{j,p} Y_{t-j}^{\mathrm{AR}}(p) + Z_t^{\mathrm{AR}}(p), \tag{3.4}$$

where

$$(a_{1,p}, \dots, a_{p,p}) := \operatorname*{argmin}_{b_{1,p},\dots,b_{p,p}} \mathbb{E}\left(Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j}\right)^2 \tag{3.5}$$

and $(Z_t^{\mathrm{AR}}(p))_{t\in\mathbb{Z}}$ is a Gaussian white noise process with mean zero and variance

$$\sigma_p^2 = \mathbb{E}\left(Y_t - \sum_{j=1}^{p} a_{j,p} Y_{t-j}\right)^2.$$

In other words, $Y_t^{\mathrm{AR}}(p)$ corresponds to the best AR(p) model which can be fitted to the process Y_t. Lemma 2.2 in Kreiss, Paparoditis and Politis [19] ensures that for growing p

$$\sum_{k=1}^{p} (1+k)|a_{k,p} - a_k| \to 0, \tag{3.6}$$

thus the process $Y_t^{\mathrm{AR}}(p)$ becomes 'close' to the process Y_t.

The bootstrap procedure now works by fitting an AR(p) model to the observed data $X_{1,T}, \dots, X_{T,T}$, where the parameter $p = p(T)$ increases with the sample size T. To be precise, we first calculate an estimator $(\hat{a}_{1,p,T}, \dots, \hat{a}_{p,p,T})$ for

$$(a_{1,p,T}, \dots, a_{p,p,T}) = \operatorname*{argmin}_{b_{1,p,T},\dots,b_{p,p,T}} \mathbb{E}\left(X_{t,T} - \sum_{j=1}^{p} b_{j,p,T} X_{t-j,T}\right)^2 \tag{3.7}$$

and then simulate a pseudo series $X^*_{1,T}, \dots, X^*_{T,T}$ according to the model

$$X^*_{t,T} = X_{t,T}; \qquad t = 1, \dots, p,$$

$$X^*_{t,T} = \sum_{j=1}^{p} \hat{a}_{j,p,T} X^*_{t-j,T} + Z^*_j; \qquad p < t \leq T.$$

Here, the quantities Z^*_j denote independent and normal distributed random variables with mean zero and variance

$$\hat{\sigma}_p^2 := \frac{1}{T-p} \sum_{t=p+1}^{T} (\hat{z}_t - \overline{z}_T)^2, \tag{3.8}$$

where $\overline{z}_T := \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{z}_t$ and

$$\hat{z}_t := X_{t,T} - \sum_{j=1}^{p} \hat{a}_{j,p,T} X_{t-j,T} \qquad \text{for } t = p+1, \dots, T,$$

thus $\hat{\sigma}_p^2$ is the standard variance estimator of the error process $\hat{z}_t$. We now define the statistic $\hat{G}^*_T(v, \omega)$ in the same way as $\hat{G}_T(v, \omega)$ where the original observations $X_{1,T}, \dots, X_{T,T}$ are replaced by the bootstrap replicates $X^*_{1,T}, \dots, X^*_{T,T}$. To assure that this procedure approximates the limiting distribution corresponding to the null hypothesis both under the null hypothesis and the alternative, we need the following technical conditions:

Assumption 3.2.

(i) $p = p(T) \in [p_{\min}(T), p_{\max}(T)]$, *where* $p_{\max}(T) \geq p_{\min}(T) \xrightarrow{T \to \infty} \infty$ *and*

$$\frac{p^3_{\max}(T)\sqrt{\log(T)}}{\sqrt{T}} = \mathrm{O}(1). \tag{3.9}$$

(ii) *The estimators for the AR parameters defined by* (3.7) *satisfy*

$$\max_{1 \leq j \leq p} |\hat{a}_{j,p,T} - a_{j,p}| = \mathrm{O}\big(\sqrt{\log(T)/T}\big), \tag{3.10}$$

uniformly with respect to $p \leq p(T)$.

(iii) *The estimate* $\hat{\sigma}_p^2$ *defined in* (3.8) *converges in probability to* $\sigma^2 > 0$.

All assumptions are rather standard in the framework of an AR(∞) bootstrap; see, for example, Kreiß [18] or Berg, Paparoditis and Politis [3]. Thanks to (3.9), assumption (3.10) is, for example, satisfied for the least squares or the Yule–Walker estimators; see Hannan and Kavalieris [16]. The latter condition is extremely important, as it implies that $X^*_{t,T}$ shows a similar behavior as the AR(p) process $Y_t^{\mathrm{AR}}(p)$ and is therefore also 'close' to Y_t in a similar sense as (3.6). Therefore, we can expect that statistics based on the bootstrap replicates behave in the same way as those based on a stationary process. Precisely, we obtain the following result which implies consistency of the bootstrap procedure described above.

Theorem 3.3. *Suppose that the assumptions of Theorem 2.1 hold and that furthermore Assumptions 3.1 and 3.2 are satisfied. Then as $T \to \infty$ we have conditionally on $X_{1,T}, \dots, X_{T,T}$*

$$\big(\hat{G}_T^*(v,\omega)\big)_{(v,\omega)\in[0,1]^2} \Rightarrow \big(\tilde{G}(v,\omega)\big)_{v\in[0,1],\omega\in[0,1]},$$

where $(\tilde{G}(v,\omega))_{(v,\omega)\in[0,1]^2}$ denotes a centered Gaussian process with covariance structure

$$\mathrm{Cov}\big(\tilde{G}(v_1,\omega_1),\tilde{G}(v_2,\omega_2)\big) = \frac{\min(v_1,v_2)-v_1v_2}{2\pi}\int_0^{\pi\min(\omega_1,\omega_2)}\bigg(\int_0^1 f(u,\lambda)\,\mathrm{d}u\bigg)^2\mathrm{d}\lambda.$$

We now obtain empirical quantiles of $\sup_{(v,\omega)\in[0,1]^2}|G(v,\omega)|$ by calculating $\hat{D}_{T,i}^* := \sup_{(v,\omega)\in[0,1]^2}|\hat{G}_{T,i}^*(v,\omega)|$ for $i=1,\dots,B$ where $\hat{G}_{T,1}^*(v,\omega),\dots,\hat{G}_{T,B}^*(v,\omega)$ are the B bootstrap replicates of $\hat{G}_T(v,\omega)$. The null hypothesis is then rejected, whenever

$$\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}\big|\hat{D}_T(v,\omega)\big| > \big(\hat{D}_T^*\big)_{T,\lfloor(1-\alpha)B\rfloor}, \tag{3.11}$$

where $(\hat{D}_T^*)_{T,1},\dots,(\hat{D}_T^*)_{T,B}$ denotes the order statistic of $\hat{D}_{T,1}^*,\dots,\hat{D}_{T,B}^*$. The test has asymptotic level α because of Theorem 3.3 and is consistent within the class of alternatives satisfying Assumptions 3.1 and 3.2. This follows, since conditionally on $X_{1,T},\dots,X_{T,T}$ each bootstrap statistic $\sup_{(v,\omega)\in[0,1]^2}|\hat{G}_T^*(v,\omega)|$ converges to a non-degenerate random variable, while $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ converges to infinity by Theorem 2.1. We finally point out that similar results can be shown for the statistic which is obtained by replacing the localised periodogram in $\hat{D}_T$ by the pre-periodogram. The technical details are omitted for the sake of brevity, but the finite sample performance of this alternative approach will be investigated in the following section as well.

4. Finite sample properties

4.1. Choosing the parameter

We first comment on how to choose the parameters N and p in concrete applications. Although the proposed method does not show much sensitivity with respect to different choices of both parameters, we select p throughout this section as the minimizer of the AIC criterion dating back to Akaike [1], which is defined by

$$\hat{p} = \underset{p}{\operatorname{argmin}} \frac{1}{T} \sum_{k=1}^{T/2} \left(\log\left(f_{\hat{\theta}(p)}(\lambda_{k,T})\right) + \frac{I_T^X(\lambda_{k,T})}{f_{\hat{\theta}(p)}(\lambda_{k,T})} \right) + p/T$$

in the context of stationary processes. See also Whittle [30] and Whittle [31]. Here, $f_{\hat{\theta}(p)}$ is the spectral density of a stationary AR(p) process with the fitted coefficients and I_T^X is the usual stationary periodogram. Therefore, we focus in the following discussion on the sensitivity analysis of the test (3.11) with respect to different choices of N, and we will see that the particular choice of that tuning parameter has typically very little influence on the outcome of the test.

4.2. Bootstrap approximation

Let us illustrate now how well the proposed bootstrap method approximates the distribution of the statistic $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ under the null hypothesis. For this purpose, we simulate observations from the stationary AR(1) model

$$X_{t,T} = 0.5X_{t-1,T} + Z_t, \qquad t = 1,\dots,T, \tag{4.1}$$

for $T = 128$. In particular, we generate 1000 versions of this process and calculate each time the test statistic $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$, both for $N = 16$ and $N = 8$. These outcomes can be used to estimate the exact distribution of the test statistic. In a next step, we choose randomly 10 series from the 1000 replications of (4.1), for which we calculate another 1000 bootstrap approximations each. Based on these bootstrap replications, we estimate the density of the corresponding bootstrap approximations of the test statistic as well. The plots comparing these densities are given in Figure 1 where the dotted line corresponds to the estimated exact density while the dashed lines show the 10 estimated densities of the bootstrap approximations.

4.3. Size and power of the test

In this section, we investigate the size and power of the test (3.11) and the analogue based on the pre-periodogram. We also compare these methods with three other tests for stationarity, which recently have been proposed in the literature. All reported results are based on 200 bootstrap replications and 1000 simulation runs under the null hypothesis while we use 500 simulation runs under the alternative. To study the approximation of the nominal level, we simulate AR(1) processes

$$X_t = \phi X_{t-1} + Z_t, \qquad t \in \mathbb{Z}, \tag{4.2}$$

and MA(1) processes

$$X_t = Z_t + \theta Z_{t-1}, \qquad t \in \mathbb{Z}, \tag{4.3}$$

for different values of the parameters ϕ and θ, where the Z_t are independent and standard normal distributed random variables throughout the whole section. The corresponding results

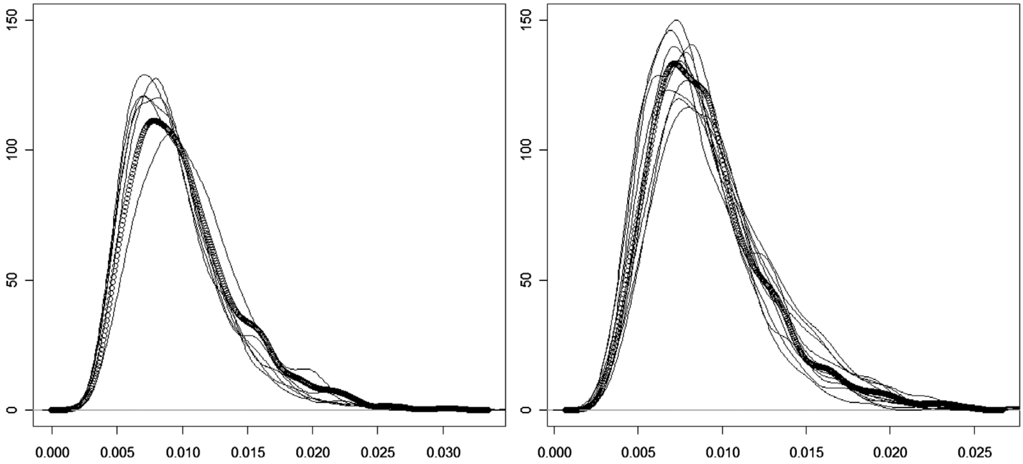

Figure 1. Estimated densities of the distribution of the statistic $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ under the null hypothesis. The dotted line is the estimated exact density while the solid lines corresponds to the estimated densities of the bootstrap approximations. Left panel: $N = 8$; right panel: $N = 16$.

are depicted in Tables 1 and 2, respectively, and we observe a precise approximation of the nominal level in the AR(1) case for $\phi \in \{-0.5, 0, 0.5, 0.9\}$ and in the MA(1) case for $\theta \in \{-0.9, -0.5, 0.5, 0.9\}$ even for very small samples sizes. Furthermore, if T gets larger, the results are basically not affected by the choice of N in these cases. For $\phi = -0.9$, the nominal level is underestimated for our choice of T, but at least if T grows the approximation of the nominal level becomes more precise.

Table 1. Rejection probabilities of the test (3.11) under the null hypothesis. The data was generated according to model (4.2)

			$\phi = -0.9$		$\phi = -0.5$		$\phi = 0$		$\phi = 0.5$		$\phi = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.021	0.069	0.025	0.060	0.035	0.086	0.050	0.099	0.044	0.108
128	16	8	0.022	0.063	0.031	0.077	0.042	0.081	0.034	0.092	0.050	0.099
128	8	16	0.020	0.066	0.030	0.076	0.038	0.083	0.055	0.102	0.038	0.081
256	32	8	0.028	0.078	0.040	0.086	0.051	0.106	0.053	0.111	0.051	0.111
256	16	16	0.016	0.063	0.038	0.089	0.044	0.085	0.045	0.080	0.033	0.085
256	8	32	0.022	0.068	0.036	0.083	0.051	0.098	0.050	0.102	0.051	0.105
512	64	8	0.020	0.073	0.054	0.103	0.052	0.084	0.042	0.090	0.039	0.112
512	32	16	0.023	0.070	0.046	0.083	0.044	0.090	0.049	0.092	0.038	0.080
512	16	32	0.029	0.067	0.038	0.079	0.056	0.098	0.052	0.099	0.048	0.101
512	8	64	0.025	0.070	0.050	0.102	0.047	0.101	0.051	0.112	0.054	0.105

Table 2. Rejection probabilities of the test (3.11) under the null hypothesis. The data was generated according to model (4.3)

			$\theta = -0.9$		$\theta = -0.5$		$\theta = 0.5$		$\theta = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.024	0.073	0.027	0.060	0.045	0.091	0.045	0.096
128	16	8	0.033	0.071	0.037	0.085	0.043	0.087	0.029	0.076
128	8	16	0.028	0.063	0.031	0.071	0.050	0.102	0.028	0.085
256	32	8	0.047	0.085	0.033	0.081	0.040	0.074	0.042	0.080
256	16	16	0.044	0.095	0.031	0.080	0.043	0.083	0.035	0.076
256	8	16	0.029	0.074	0.034	0.081	0.059	0.112	0.038	0.076
512	64	8	0.038	0.084	0.041	0.087	0.052	0.106	0.041	0.089
512	32	16	0.047	0.091	0.043	0.073	0.047	0.094	0.050	0.100
512	16	32	0.036	0.085	0.044	0.082	0.050	0.093	0.050	0.087
512	8	64	0.051	0.094	0.040	0.078	0.070	0.116	0.037	0.080

To study the power of the test (3.11), we simulate data from the following four models which all correspond to the alternative of non-stationary processes. In particular, we consider

$$X_{t,T} = (1 + t/T)Z_t, \tag{4.4}$$

$$X_{t,T} = -0.9\sqrt{\frac{t}{T}}X_{t-1,T} + Z_t, \tag{4.5}$$

$$X_{t,T} = \begin{cases} 0.5X_{t-1} + Z_t, & \text{if } 1 \le t \le \dfrac{T}{2}, \\ -0.5X_{t-1} + Z_t, & \text{if } \dfrac{T}{2} + 1 \le t \le T, \end{cases} \tag{4.6}$$

$$X_{t,T} = Z_t + 0.8\cos\bigl(1.5 - \cos(4\pi t/T)\bigr)Z_{t-q}, \tag{4.7}$$

where we display the results for the last model for different $q \in \mathbb{N}$. Note that due to Remark 2.5 the alternative (4.6) also fits into the theoretical framework. The corresponding rejection probabilities are reported in Table 3 and we observe a reasonable behavior of the procedure in the first three considered cases, whereas power is rather low for the alternative (4.7). Similar to the null hypothesis we observe robustness with respect to different choices of N, and even for the choice $M = 32$, $N = 8$, which appears to be implausible in view of (2.8), the results are satisfying. It might be of interest to compare these results both with the pre-periodogram approach from Theorem 2.2 and with other tests for the hypothesis of stationarity which have been recently suggested in the literature. In particular, we consider the tests of Paparoditis [23], Dwivedi and Subba Rao [15] and Dette, Preuss and Vetter [14].

In Table 4, we present the rejection frequencies for the test based on the pre-periodogram as defined in (2.13). Recall that the use of the pre-periodogram does not require the specification of the value N, which specifies the number of observations for the calculation of the local periodogram. This makes its use attractive for practitioners. However, the results of the simulation

Table 3. Rejection probabilities of the test (3.11) for several alternatives

			(4.4)		(4.5)		(4.6)		(4.7) $q = 1$		(4.7) $q = 6$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.286	0.444	0.186	0.328	0.168	0.270	0.046	0.098	0.052	0.104
128	16	8	0.686	0.772	0.396	0.546	0.308	0.466	0.090	0.154	0.072	0.130
128	8	16	0.624	0.758	0.382	0.578	0.410	0.548	0.082	0.144	0.080	0.136
256	32	8	0.958	0.974	0.672	0.814	0.742	0.912	0.110	0.186	0.102	0.166
256	16	16	0.942	0.978	0.698	0.814	0.640	0.806	0.118	0.202	0.098	0.166
256	8	32	0.944	0.970	0.760	0.868	0.672	0.808	0.118	0.210	0.086	0.144

study show that compared to the local periodogram the use of the pre-periodogram yields to a substantial loss of power for all four alternatives. In particular for alternatives of the form (4.6), the test cannot be recommended.

In Table 5, we show the corresponding rejection probabilities for the test proposed in Dette, Preuss and Vetter [14], which is the only of the remaining methods depending on one regularization parameter only. These authors proposed to estimate the L_2 distance

$$\int_0^1 \int_0^\pi \left(f(u,\lambda) - \int_0^1 f(v,\lambda)\,\mathrm{d}v \right)^2 \mathrm{d}\lambda\,\mathrm{d}u$$

using sums of the (squared) periodogram. In order to provide a fair comparison between the two methods, we also employ the AR(∞) bootstrap to the corresponding test to generate critical values. It turns out that without a bootstrap the method of Dette, Preuss and Vetter [14] is much more sensitive with respect to different choices of N. We observe that the new method also outperforms the test proposed by Dette, Preuss and Vetter [14] in the alternatives (4.4) and (4.5). In most cases the differences are substantial. On the other hand, for the alternative (4.6) the procedure of Dette, Preuss and Vetter [14] has larger power if $T = 64$ and $T = 128$, but for $T = 256$ the novel method performs better in this case as well. Nevertheless, the new approach is clearly outperformed by the proposal of Dette, Preuss and Vetter [14] for the alternative (4.7).

Table 4. Rejection probabilities of the test based on the pre-periodogram for several alternatives

	(4.4)		(4.5)		(4.6)		(4.7) $q = 1$		(4.7) $q = 6$	
T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	0.188	0.340	0.080	0.202	0.022	0.056	0.024	0.076	0.044	0.102
128	0.552	0.702	0.216	0.392	0.036	0.116	0.038	0.086	0.052	0.098
256	0.938	0.968	0.580	0.734	0.080	0.176	0.062	0.150	0.088	0.132

Table 5. Rejection probabilities of the test proposed by Dette, Preuss and Vetter [14] for several alternatives (quantiles obtained by AR(∞) bootstrap)

			(4.4)		(4.5)		(4.6)		(4.7) $q = 1$		(4.7) $q = 6$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.116	0.196	0.188	0.232	0.250	0.344	0.244	0.350	0.056	0.116
128	16	8	0.106	0.160	0.256	0.330	0.370	0.552	0.490	0.584	0.226	0.336
128	8	16	0.168	0.268	0.220	0.286	0.432	0.566	0.398	0.516	0.072	0.126
256	32	8	0.378	0.498	0.282	0.412	0.746	0.922	0.740	0.836	0.532	0.670
256	16	16	0.208	0.368	0.276	0.410	0.618	0.794	0.716	0.816	0.342	0.444
256	8	32	0.224	0.338	0.300	0.418	0.582	0.744	0.620	0.760	0.104	0.178

In Table 6, we show the rejection frequencies for the method which was proposed in Paparoditis [23]. This concept basically works by estimating

$$\sup_{v\in[0,1]} \int_{-\pi}^{\pi} \left(\frac{f(v,\lambda)}{\int_0^1 f(u,\lambda)\,\mathrm{d}u} - 1 \right)^2 \mathrm{d}\lambda$$

via a smoothed local periodogram, which requires the choice of a smoothing bandwidth besides the window length N. We choose the uniform kernel function, and as recommended by the author we select the bandwidth via the cross validation criterion of Beltrão and Bloomfield [2]. To provide a fair comparison, we also use the AR(∞) bootstrap to obtain critical values. For the alternatives (4.4)–(4.6) the proposal of Paparoditis [23] yields substantial less power than the approach proposed in this paper, whereas for the alternative (4.7) no clear picture can be drawn. For $q = 1$, the method of Paparoditis [23] performs better, while there is no significant difference in the performance if $q = 6$. In any case, Paparoditis [23] is clearly outperformed by the approach of Dette, Preuss and Vetter [14] for (4.7).

Table 6. Rejection probabilities of the test proposed by Paparoditis [23] for several alternatives (quantiles obtained by AR(∞) bootstrap)

			(4.4)		(4.5)		(4.6)		(4.7) $q = 1$		(4.7) $q = 6$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.054	0.126	0.050	0.122	0.078	0.170	0.058	0.104	0.034	0.064
128	16	8	0.150	0.242	0.158	0.262	0.112	0.198	0.128	0.218	0.082	0.140
128	8	16	0.066	0.154	0.120	0.254	0.166	0.270	0.080	0.170	0.034	0.066
256	32	8	0.304	0.424	0.248	0.380	0.298	0.448	0.288	0.428	0.102	0.180
256	16	16	0.234	0.344	0.276	0.404	0.258	0.374	0.288	0.420	0.120	0.174
256	8	32	0.126	0.226	0.240	0.374	0.298	0.376	0.158	0.266	0.050	0.106

Table 7. Rejection probabilities of the test proposed by Dwivedi and Subba Rao [15] for several alternatives (quantiles obtained by AR(∞) bootstrap)

	(4.4)		(4.5)		(4.6)		(4.7) $q=1$		(4.7) $q=6$	
T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
64	0.174	0.266	0.056	0.100	0.082	0.164	0.072	0.120	0.046	0.098
128	0.274	0.386	0.058	0.114	0.122	0.208	0.126	0.206	0.092	0.162
256	0.604	0.716	0.128	0.210	0.174	0.276	0.234	0.340	0.174	0.272

Finally, we compare our approach to that proposed in Dwivedi and Subba Rao [15]. These authors suggested a Portmanteau type test by estimating

$$T\sum_{r=1}^{m}\left|c_T(r)\right|^2,$$

where $c_T(r)$ is the covariance of the process at lag r. For the estimation of $c_T(r)$, the authors require the choice of a smoothing bandwidth, and again we use the cross validation criterion and the uniform kernel function. Dwivedi and Subba Rao [15] also have to choose the maximal lag $m \in N$ up to which they want to estimate $c_T(r)$, and we pick $m = 5$ in the simulations. As in the other examples, we employ the AR(∞) bootstrap, and the results are given in Table 7. A comparison with our method yields a result similar to the approach of Paparoditis [23]. Our approach performs better for the alternatives (4.4)–(4.6) while the proposal of Dwivedi and Subba Rao [15] yields a higher power in model (4.7). Again it is clearly outperformed in this case by the test proposed in Dette, Preuss and Vetter [14].

4.4. Data example

As an illustration, we consider $T = 257$ observations of weekly egg prices at a German agriculture market between April 1967 and March 1972. A plot of the data is given in Figure 2, and following Paparoditis [23] the first order difference $\Delta_t = X_t - X_{t-1}$ of the observed time series are analyzed. Although several stationary models were proposed in the literature to fit this data (cf. Paparoditis [23]), the new test rejects the null hypothesis with p-value 0.006 if we choose $N = 32$ or $N = 16$, and with p-value 0.001 if we choose $N = 8$. These results demonstrate again that the choice of N does not have too much influence on the outcome, and that even the somewhat implausible choice of $N = 8$ yields a p-value similar to the others.

Note that in Paparoditis [23] a longer version of the above time series was analyzed, namely 1201 observations of weekly egg prices between April 1967 and May 1990. However, we obtain a p-value of exactly 0 even if we choose 10^6 bootstrap replicates in this case, which is why we consider the first 257 datapoints only. Paparoditis [23] rejects the null hypothesis of stationarity at level 5% if the whole dataset is used, but his approach yields a p-value of 0.1834 if it is applied to the first 257 observations of the time series only, and therefore the hypothesis of stationarity

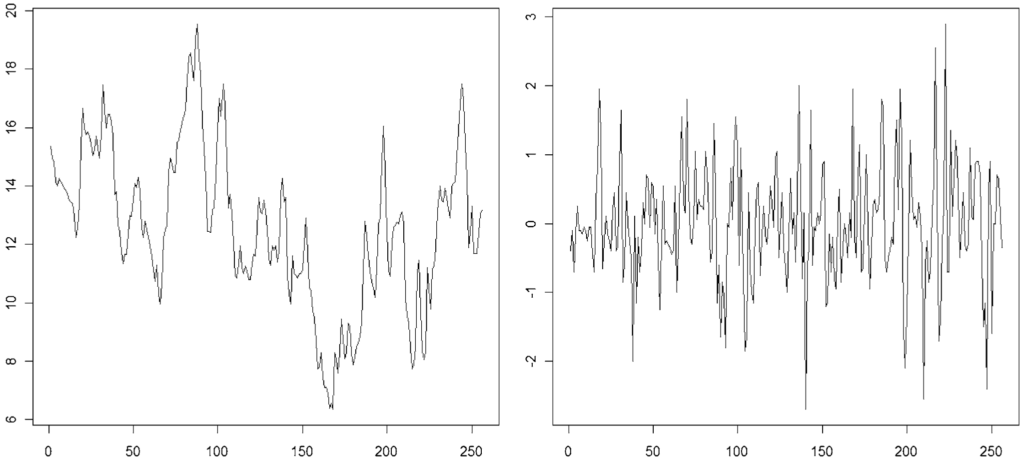

Figure 2. Left panel: Weekly egg prices at a German agriculture market between April 1967 and March 1972. Right panel: First order difference of the weekly egg prices.

cannot be rejected at a reasonable size using his method. Roughly the same p-value, namely 0.189, can be observed if the approach of Dwivedi and Subba Rao [15] is employed.

Appendix: Proofs

A.1. Proof of Theorem 2.1

Throughout the proof, we set $y_j = (v_j, \omega_j) \in [0,1]^2$ for $j = 1, \dots, K$ and $K \in \mathbb{N}$. To show weak convergence we follow Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner [28] and prove the following two claims:

(1) Convergence of the finite dimensional distributions, that is,

$$\big(\hat{G}_T(y_j)\big)_{j=1,\dots,K} \xrightarrow{D} \big(G(y_j)\big)_{j=1,\dots,K}. \tag{A.1}$$

(2) Stochastic equicontinuity, that is,

$$\forall \eta, \varepsilon > 0 \ \exists \delta > 0: \qquad \lim_{T\to\infty} P\Big(\sup_{y_1,y_2\in[0,1]^2 \,:\, d_2(y_1,y_2)<\delta} \big|G_T(y_1) - G_T(y_2)\big| > \eta\Big) < \varepsilon, \tag{A.2}$$

where $d_2(y_1, y_2) = \sqrt{(v_1 - v_2)^2 + (w_1 - w_2)^2}$.

Proof of (A.1). The claim follows from similar arguments as given in the proof of Theorem 3.1 in Dette, Preuss and Vetter [14]. For the sake of brevity and because we will use similar arguments in the proof of (A.2), we will sketch how the assertions

$$\mathbb{E}\big(\hat{G}_T(v,\omega)\big) \xrightarrow{T\to\infty} 0, \tag{A.3}$$

$$\begin{aligned}\operatorname{Cov}\big(\hat{G}_T(y_1),\hat{G}_T(y_2)\big) \xrightarrow{T\to\infty} \frac{1}{2\pi}\int_0^1\int_0^{\pi\min(\omega_1,\omega_2)} &\big(1_{[0,v_1]}(u)-v_1\big)\big(1_{[0,v_2]}(u)-v_2\big) \\ &\times f^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\end{aligned} \tag{A.4}$$

can be shown. Note that we have

$$\begin{aligned}\hat{G}_T(v,\omega) &= \frac{1}{\sqrt{T}}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)I_N^X(u_j,\lambda_k) - \sqrt{T}D_{N,M}(\phi_{v,\omega,M,N}) \\ &=: G_T(\phi_{v,\omega,M,N})\end{aligned}$$

with

$$\phi_{v,\omega,M,N}(u,\lambda) := \left(I_{[0,\lfloor vM\rfloor/M]}(u) - \frac{\lfloor vM\rfloor}{M}\right)I_{[0,2\pi\lfloor\omega N/2\rfloor/N]}(\lambda)$$

for $u,\lambda \ge 0$ and

$$D_{N,M}(\phi) := \frac{1}{2\pi}\int_0^1\int_0^{\pi}\phi(u,\lambda)f(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u.$$

In order to simplify some technical arguments, we also define

$$\phi_{v,\omega,M,N}(u,\lambda) := \phi_{v,\omega,M,N}(u,-\lambda)$$

for $u \ge 0, \lambda < 0$ and obtain from (2.2)

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)I_N^X(u_j,\lambda_k)\right) \\ &\quad= \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k) \\ &\qquad\qquad\times\frac{1}{2\pi N}\sum_{p,q=0}^{N-1}\sum_{l,m=-\infty}^{\infty}\psi_l\left(\frac{t_j-N/2+1+p}{T}\right)\psi_m\left(\frac{t_j-N/2+1+q}{T}\right) \\ &\qquad\qquad\qquad\times\mathbb{E}(Z_{t_j-N/2+1+p-m}Z_{t_j-N/2+1+q-l}) \\ &\qquad\qquad\qquad\times\exp\big(-\mathrm{i}\lambda_k(p-q)\big)\big(1+\mathrm{O}(1/T)\big).\end{aligned}$$

A Taylor expansion now yields that this term becomes

$$
\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)\frac{1}{2\pi N}\\
&\qquad\times\sum_{p,q=0}^{N-1}\sum_{l,m=-\infty}^{\infty}\psi_l(u_j)\psi_m(u_j)\\
&\qquad\qquad\times\mathbb{E}(Z_{t_j-N/2+1+p-m}Z_{t_j-N/2+1+q-l})\\
&\qquad\qquad\times\exp\bigl(-\mathrm{i}\lambda_k(p-q)\bigr)\bigl(1+\mathrm{O}(1/T)+\mathrm{O}\bigl(N^2/T^2\bigr)\bigr).
\end{aligned}
$$

See Dette, Preuss and Vetter [14] for details. Since $\mathbb{E}(Z_i Z_j)=0$ for $i\neq j$, we obtain the equation $p=q+m-l$ which shows that the above expression equals

$$
\begin{aligned}
&\frac{1}{2\pi NT}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)\\
&\qquad\times\sum_{l,m=-\infty}^{\infty}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}\psi_l(u_j)\psi_m(u_j)\exp\bigl(-\mathrm{i}\lambda_k(m-l)\bigr)\\
&\quad+\mathrm{O}(1/T)+\mathrm{O}\bigl(N^2/T^2\bigr)\\
&=\frac{1}{2\pi NT}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)\\
&\qquad\times\sum_{\substack{l,m=-\infty\\|l-m|\le N-1}}^{\infty}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}\psi_l(u_j)\psi_m(u_j)\exp\bigl(-\mathrm{i}\lambda_k(m-l)\bigr)\\
&\quad+\frac{1}{2\pi NT}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,M,N}(u_j,\lambda_k)\\
&\qquad\times\sum_{\substack{l,m=-\infty\\|l-m|\ge N}}^{\infty}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}\psi_l(u_j)\psi_m(u_j)\exp\bigl(-\mathrm{i}\lambda_k(m-l)\bigr)\\
&\quad+\mathrm{O}(1/T)+\mathrm{O}\bigl(N^2/T^2\bigr).
\end{aligned}
$$

Dropping the extra condition $0 \le q + m - l \le N - 1$, the second term is bounded by

$$
\begin{aligned}
C \sum_{\substack{l,m=-\infty \\ |l-m| \ge N}}^{\infty} \sup_u \bigl|\psi_l(u)\bigr| \sup_u \bigl|\psi_m(u)\bigr| &\le 2C \sum_{m=-\infty}^{\infty} \sup_u \bigl|\psi_m(u)\bigr| \sum_{\substack{l=-\infty \\ |l| \ge N/2}}^{\infty} \sup_u \bigl|\psi_l(u)\bigr| \\
&\le \frac{4C \sum_{m=-\infty}^{\infty} \sup_u |\psi_m(u)| \sum_{l=-\infty}^{\infty} |l| \sup_u |\psi_l(u)|}{N} \qquad \text{(A.5)} \\
&= \mathrm{O}(1/N)
\end{aligned}
$$

for some $C \in \mathbb{R}$ and the order follows from (2.3). Using (2.3) and (A.5) in the same way again, the first quantity above can be shown to be equal to

$$
\frac{1}{2\pi T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,M,N}(u_j, \lambda_k) \sum_{l,m=-\infty}^{\infty} \psi_l(u_j) \psi_m(u_j) \exp\bigl(-\mathrm{i}\lambda_k(m-l)\bigr) + \mathrm{O}(1/N),
$$

and therefore we obtain

$$
\begin{aligned}
&\mathbb{E}\Biggl(\frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,M,N}(u_j, \lambda_k) I_N^X(u_j, \lambda_k)\Biggr) \\
&\quad = \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,M,N}(u_j, \lambda_k) f(u_j, \lambda_k) + \mathrm{O}(1/N) \\
&\qquad + \mathrm{O}\bigl(N^2/T^2\bigr) + \mathrm{O}(1/T) \\
&\quad = D_{N,M}(\phi_{v,\omega,M,N}) + \mathrm{O}(1/N) + \mathrm{O}\bigl(N^2/T^2\bigr) + \mathrm{O}(1/T),
\end{aligned}
$$

where the order of the Riemann approximation follows from the specific choice of the mid-points u_j. This together with (2.8) yields (A.3).

To prove (A.4), we use symmetry arguments and obtain

$$
\begin{aligned}
&T \operatorname{cum}\Biggl(\frac{1}{T} \sum_{j_1=1}^{M} \sum_{k_1=1}^{N/2} \phi_{v_1,\omega_1,M,N}(u_{j_1}, \lambda_{k_1}) I_N^X(u_{j_1}, \lambda_{k_1}), \\
&\qquad \frac{1}{T} \sum_{j_2=1}^{M} \sum_{k_2=1}^{N/2} \phi_{v_2,\omega_2,M,N}(u_{j_2}, \lambda_{k_2}) I_N^X(u_{j_2}, \lambda_{k_2})\Biggr) \\
&\quad = \frac{1}{4T} \frac{1}{(2\pi N)^2} \sum_{j_1,j_2=1}^{M} \sum_{k_1,k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi_{v_1,\omega_1,M,N}(u_{j_1}, \lambda_{k_1}) \phi_{v_2,\omega_2,M,N}(u_{j_2}, \lambda_{k_2})
\end{aligned}
$$

$$\times \sum_{p_1,p_2,q_1,q_2=0}^{N-1} \sum_{m_1,m_2,l_1,l_2=-\infty}^{\infty} \psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})$$
$$\times \exp\bigl(-\mathrm{i}\lambda_{k_1}(p_1-q_1)\bigr)\exp\bigl(-\mathrm{i}\lambda_{k_2}(p_2-q_2)\bigr)$$
$$\times \operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_1}-N/2+1+q_1-l_1}, Z_{t_{j_2}-N/2+1+p_2-m_2} Z_{t_{j_2}-N/2+1+q_2-l_2})$$
$$\times \bigl(1+\mathrm{O}\bigl(N^2/T^2\bigr)+\mathrm{O}(1/T)\bigr)$$

in the same way as above. Because of

$$\begin{aligned}
&\operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_1}-N/2+1+q_1-l_1}, Z_{t_{j_2}-N/2+1+p_2-m_2} Z_{t_{j_2}-N/2+1+q_2-l_2})\\
&\quad= \operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_2}-N/2+1+q_2-l_2}) \operatorname{cum}(Z_{t_{j_2}-N/2+1+p_2-m_2} Z_{t_{j_1}-N/2+1+q_1-l_1})\\
&\qquad+ \operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_2}-N/2+1+p_2-m_2}) \operatorname{cum}(Z_{t_{j_1}-N/2+1+q_1-l_1} Z_{t_{j_2}-N/2+1+q_2-l_2}),
\end{aligned}$$

the calculation of the highest order term in the variance splits into two sums and we only consider the first one (the second sum is treated completely analogously), which equals

$$\begin{aligned}
&\frac{1}{4T}\sum_{j_1,j_2=1}^{M}\sum_{k_1,k_2=-\lfloor (N-1)/2\rfloor}^{N/2} \phi_{v_1,\omega_1,M,N}(u_{j_1},\lambda_{k_1})\phi_{v_2,\omega_2,M,N}(u_{j_2},\lambda_{k_2})\frac{1}{(2\pi N)^2}\\
&\quad\times \sum_{m_1,m_2,l_1,l_2=-\infty}^{\infty} \sum_{\substack{q_1,q_2=0\\ 0\le q_2+m_1-l_2+t_{j_2}-t_{j_1}\le N-1\\ 0\le q_1+m_2-l_1+t_{j_1}-t_{j_2}\le N-1}}^{N-1} \psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})\\
&\quad\times \exp\bigl(-\mathrm{i}(\lambda_{k_1}-\lambda_{k_2})(q_2-q_1+t_{j_2}-t_{j_1})\bigr)\\
&\quad\times \exp\bigl(-\mathrm{i}\lambda_{k_1}(m_1-l_2)-\mathrm{i}\lambda_{k_2}(m_2-l_1)\bigr)\\
&= \frac{1}{4T}\sum_{j_1,j_2=1}^{M}\sum_{k_1,k_2=-\lfloor (N-1)/2\rfloor}^{N/2} \phi_{v_1,\omega_1,M,N}(u_{j_1},\lambda_{k_1})\phi_{v_2,\omega_2,M,N}(u_{j_2},\lambda_{k_2})\frac{1}{(2\pi N)^2}\\
&\quad\times \sum_{\substack{m_1,m_2,l_1,l_2=-\infty\\ (+)}}^{\infty} \sum_{\substack{q_1,q_2=0\\ 0\le q_2+m_1-l_2+t_{j_2}-t_{j_1}\le N-1\\ 0\le q_1+m_2-l_1+t_{j_1}-t_{j_2}\le N-1}}^{N-1} \psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})\\
&\quad\times \exp\bigl(-\mathrm{i}(\lambda_{k_1}-\lambda_{k_2})(q_2-q_1+t_{j_2}-t_{j_1})\bigr)\\
&\quad\times \exp\bigl(-\mathrm{i}\lambda_{k_1}(m_1-l_2)-\mathrm{i}\lambda_{k_2}(m_2-l_1)\bigr)\\
&\quad\times \bigl(1+\mathrm{O}(1/N)\bigr),
\end{aligned}$$

where $\sum_{(+)}$ means that summation is only performed over those indices $x, y \in \{m_1, m_2, l_1, l_2\}$ such that $|x - y| < N$, and the $O(1/N)$-term follows with (A.5). Assume that j_1 has been chosen. Then j_2 must be equal to j_1, $j_1 - 1$ or $j_1 + 1$, as all other combination of j_1 and j_2 vanish, because of the condition $0 \le q_2 + m_1 - l_2 + t_{j_2} - t_{j_1} \le N - 1$ and the fact that the summation is only performed with respect to the indices satisfying $|x - y| < N$. If j_2 equals $j_1 - 1$ or $j_1 + 1$, it follows from the conditions on q_1 and q_2 that for chosen m_i and l_i, there are at most $|m_2 - l_1|$ possible choices for q_1 and at most $|m_1 - l_2|$ possible choices for q_2. It therefore follows with (2.3) that the terms with $j_2 \in \{j_1 - 1, j_1 + 1\}$ are of order $O(1/N)$.

Therefore, we only have to consider the case $j_1 = j_2$, and the above expression is

$$
\begin{aligned}
&\frac{1}{4T} \sum_{j_1=1}^{M} \sum_{k_1,k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi_{v_1,\omega_1,M,N}(u_{j_1}, \lambda_{k_1}) \phi_{v_2,\omega_2,M,N}(u_{j_1}, \lambda_{k_2}) \\
&\quad \times \frac{1}{(2\pi N)^2} \sum_{\substack{m_1,m_2,l_1,l_2=-\infty \\ (+)}}^{\infty} \sum_{\substack{q_1,q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1 \\ 0 \le q_1+m_2-l_1 \le N-1}}^{N-1} \psi_{m_1}(u_{j_1}) \psi_{l_1}(u_{j_1}) \psi_{m_2}(u_{j_2}) \psi_{l_2}(u_{j_2}) \\
&\qquad \times \exp\bigl(-i(\lambda_{k_1} - \lambda_{k_2})(q_2 - q_1)\bigr) \\
&\qquad \times \exp\bigl(-i\lambda_{k_1}(m_1 - l_2) - i\lambda_{k_2}(m_2 - l_1)\bigr) \\
&\qquad \times \bigl(1 + O(1/N)\bigr).
\end{aligned}
\tag{A.6}
$$

Observing

$$
\frac{1}{N} \sum_{q=0}^{N-1} \exp\bigl(-i(\lambda_{k_1} - \lambda_{k_2})q\bigr) = \begin{cases} 1, & k_1 - k_2 = lN \text{ with } l \in \mathbb{Z}, \\ 0, & \text{else}, \end{cases}
$$

it follows that for fixed m_1, l_2 and $k_1 \ne k_2$ we have

$$
\begin{aligned}
\Biggl| \sum_{\substack{q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1}}^{N-1} \exp\bigl(-i(\lambda_{k_1} - \lambda_{k_2})q_2\bigr) \Biggr| &= \Biggl| \sum_{\substack{q_2=0 \\ q_2+m_1-l_2<0 \\ \text{or} \\ q_2+m_1-l_2>N-1}}^{N-1} \exp\bigl(-i(\lambda_{k_1} - \lambda_{k_2})q_2\bigr) \Biggr| \\
&\le |m_1 - l_2|,
\end{aligned}
$$

which implies

$$
\Biggl| \frac{1}{(2\pi N)^2} \sum_{\substack{q_1,q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1 \\ 0 \le q_1+m_2-l_1 \le N-1}}^{N-1} \exp\bigl(-i(\lambda_{k_1} - \lambda_{k_2})(q_2 - q_1)\bigr) \Biggr| \le |m_1 - l_2||m_2 - l_1|/(2\pi N)^2. \tag{A.7}
$$

By using (2.3) and (A.7), it can now be seen that all terms with $k_1 \neq k_2$ are of the order $\mathrm{O}(1/N)$, and similar arguments as used in the calculation of the expectation yield that (A.6) equals

$$\frac{1}{4\pi}\int_0^1\int_0^{\pi\min(\omega_1,\omega_2)}\big(1_{[0,v_1]}(u)-v_1\big)\big(1_{[0,v_2]}(u)-v_2\big)f^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u$$
$$+\mathrm{O}(1/N)+\mathrm{O}\big(N^2/T^2\big). \qquad \square$$

Proof of (A.2). Note that

$$\mathcal{F}_T := \big\{\phi_{v,\omega,M,N};\, v,\omega\in[0,1]\big\} = \big\{\phi_{v,\omega,M,N};\,(v,\omega)\in P_T\big\},$$

where

$$P_T := \left\{0,\frac{1}{M},\frac{2}{M},\dots,\frac{M-1}{M},1\right\}\times\left\{0,\frac{2}{N},\frac{4}{N},\dots,1-\frac{2}{N},1\right\}$$

(recall that N is assumed to be even throughout this paper). We define

$$\rho_2(\phi) := \left(\int_0^1\int_0^{\pi}\phi^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\right)^{1/2},$$

and $\mathcal{F}_T^2$ is the set of functions, which can be expressed as a sum or a difference of two elements in $\mathcal{F}_T$. The main task is to prove the following theorem.

Theorem A.1. *There exists a constant $C\in\mathbb{R}$ such that for all $\phi\in\mathcal{F}_T^2$:*

$$\mathbb{E}\big(\big|\hat{G}_T(\phi)\big|^k\big) \le (2k)!C^k\rho_2(\phi)^k \qquad \forall k\in\mathbb{N}\ \textit{even}.$$

Stochastic equicontinuity follows then by similar arguments as given in Dahlhaus [6], which is why we will only sketch the main steps and refer to his work for most details. The first consequence of Theorem A.1 regards the existence of a constant $C_1\in\mathbb{R}$ such that for all $g,h\in\mathcal{F}_T$ and $\eta>0$:

$$P\big(\big|\hat{G}_T(g)-\hat{G}_T(h)\big|>\eta\rho_2(g-h)\big)\le 96\exp\left(-\sqrt{\frac{\eta}{C_1}}\right).$$

A straightforward modification of the chaining lemma in Chapter VII.2 of Pollard [24] then yields that for a stochastic process $(Z(v))_{v\in V}$, whose index set V has a finite covering integral

$$J(\delta)=\int_0^{\delta}\left[\log\left(\frac{48N(u)^2}{u}\right)\right]^2\mathrm{d}u \tag{A.8}$$

for all δ and which satisfies

$$P\big(\big|Z(v)-Z(w)\big|>\nu d(v,w)\big)\le 96\exp\left(-\sqrt{\frac{\nu}{C_1}}\right)$$

for a semi-metric d on V and a constant $C_1 \in \mathbb{R}$, there exist a dense subset $V^* \subset V$ such that

$$P\big(\exists v, w \in V^* \text{ with } d(v,w) < \varepsilon \text{ and } \big|Z(v) - Z(w)\big| > 26C_1 J\big(d(v,w)\big)\big) \leq 2\varepsilon.$$

In (A.8), $N(u)$ is the covering number which is defined as the smallest number $m \in \mathbb{N}$ for which there exist $z_1, \dots, z_m \in V$ with $\min_i d(z, z_i) \leq u$ for all $z \in V$. By using $y_i = (v_i, \omega_i)$, we obtain

$$\begin{aligned} &P\Big(\sup_{y_1, y_2 \in P_T : d_2(y_1, y_2) < \delta} \big|\hat{G}_T(v_2, w_2) - \hat{G}_T(v_1, w_1)\big| > \eta\Big) \\ &\quad \leq P\Big(\sup_{f,g \in \mathcal{F}_T : \rho_2(f,g) < \varepsilon(\delta)} \big|\hat{G}_T(f) - \hat{G}_T(g)\big| > \eta\Big) \end{aligned}$$

for a certain sequence $\varepsilon(\delta) \xrightarrow{\delta \to 0} 0$ by continuity. The right-hand side of this inequality equals

$$\begin{aligned} &P\Big(\sup_{f,g \in \mathcal{F}_T : \rho_2(f,g) < \varepsilon(\delta)} \big|\hat{G}_T(f) - \hat{G}_T(g)\big| > \eta, \eta \geq 26C_1 J_T\big(\varepsilon(\delta)\big)\Big) \\ &\qquad + P\Big(\sup_{f,g \in \mathcal{F}_T : \rho_2(f,g) < \varepsilon(\delta)} \big|\hat{G}_T(f) - \hat{G}_T(g)\big| > \eta, \eta < 26C_1 J_T\big(\varepsilon(\delta)\big)\Big) \\ &\quad \leq 2\varepsilon(\delta) + P\big(\eta < 26C_1 J_T\big(\varepsilon(\delta)\big)\big), \end{aligned}$$

where $J_T(\delta)$ is the corresponding covering integral of $\mathcal{F}_T$. Note that $\eta < 26C_1 J_T(\varepsilon(\delta))$ is not random and that $J_T(\delta)$ can be bounded by $J(\delta)$, which is the covering integral of $\bigcup_{i=1}^{\infty} \mathcal{F}_i$ (which is finite for every δ). Because of $J(\varepsilon(\delta)) \xrightarrow{\delta \to 0} 0$, we have $\eta > 26C_1 J(\delta)$ whenever δ is sufficiently small and obtain

$$P\Big(\sup_{f,g \in \mathcal{F}_T : \rho_2(f,g) < \varepsilon(\delta)} \big|\hat{G}_T(f) - \hat{G}_T(g)\big| > \eta\Big) < 2\varepsilon(\delta),$$

which implies the stochastic equicontinuity.

Proof of Theorem A.1. We show

$$\big|\mathrm{cum}_l\big(\sqrt{T}\hat{D}_T(\phi)\big)\big| \leq (2l)!\tilde{C}^l \rho_2(\phi)^l \qquad \forall l \in \mathbb{N}, \tag{A.9}$$

where

$$\hat{D}_T(\phi) := \frac{1}{\sqrt{T}}\hat{G}_T(\phi) + D_{N,M}(\phi).$$

Since $D_{N,M}(\phi)$ is constant, this implies

$$\big|\mathrm{cum}_l(\hat{G}_T)\big| \leq (2l)!C^l \rho_2(\phi)^l \qquad \forall l \in \mathbb{N}$$

for some C, and then it follows as in Dahlhaus [6] that

$$\begin{aligned}
\mathbb{E}\big(\big|\hat{G}_T(\phi)\big|^k\big) &= \Bigg|\sum_{\substack{\{P_1,\dots,P_m\}\\ \text{Partition of}\\ \{1,\dots,k\}}} \Bigg\{\prod_{j=1}^{m} \operatorname{cum}_{|P_j|}\big(\hat{G}_T(\phi)\big)\Bigg\}\Bigg| \\
&\le \rho_2(\phi)^k C^k \sum_{\substack{\{P_1,\dots,P_m\}\\ \text{Partition of}\\ \{1,\dots,k\}}} \prod_{j=1}^{m} \big(2|P_j|\big)! \\
&\le (2k)! C^k 2^k \rho_2(\phi)^k,
\end{aligned}$$

since we only consider the case where k is even. This yields the assertion.

In order to prove (A.9), we assume without loss of generality that l is even, as the case for odd l is proved in the same way. The lth cumulant of $\sqrt{T}\hat{D}_T(\phi)$ is given by

$$\begin{aligned}
&\frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2} \phi(u_{j_1},\lambda_{k_1})\cdots\phi(u_{j_l},\lambda_{k_l}) \frac{1}{(2\pi N)^l} \\
&\quad\times \sum_{p_1,q_1,p_2,\dots,p_l,q_l=0}^{N-1} \sum_{m_1,n_1,m_2,\dots,m_l,n_l=-\infty}^{\infty} \psi_{m_1}(u_{j_1})\cdots\psi_{n_l}(u_{j_l}) \\
&\qquad\times \operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_1}-N/2+1+q_1-n_1},\dots, \\
&\qquad\qquad Z_{t_{j_l}-N/2+1+p_l-m_l} Z_{t_{j_l}-N/2+1+q_l-n_l}) \\
&\qquad\times \exp\big(-\mathrm{i}\lambda_{k_1}(p_1-q_1)\big)\cdots\exp\big(-\mathrm{i}\lambda_{k_l}(p_l-q_l)\big) \\
&\qquad\times \big(1+\mathrm{O}\big(N^2/T^2\big)+\mathrm{O}(1/T)\big),
\end{aligned}$$

where both $\mathrm{O}(\cdot)$-terms follow as in the proof of (A.3). We define $Y_{i,1} := Z_{t_{j_i}-N/2+1+p_i-m_i}$ and $Y_{i,2} := Z_{t_{j_i}-N/2+1+q_i-n_i}$ for $i \in \{1,\dots,l\}$. Theorem 2.3.2 in Brillinger [4] yields

$$\operatorname{cum}_l\big(\sqrt{T}\hat{D}_T(\phi)\big) = \sum_{\nu} V_T(\nu)\big(1+\mathrm{O}\big(N^2/T^2\big)+\mathrm{O}(1/T)\big),$$

where the sum runs over all indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_l$ with $|\nu_i| = 2$ ($1 \le i \le l$, due to Gaussianity) of the matrix

$$\begin{matrix} Y_{1,1} & Y_{1,2} \\ \vdots & \vdots \\ Y_{l,1} & Y_{l,2} \end{matrix} \tag{A.10}$$

and

$$V_T(\nu) := \frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2} \phi(u_{j_1}, \lambda_{k_1}) \cdots \phi(u_{j_l}, \lambda_{k_l})$$
$$\times \frac{1}{(2\pi N)^l} \sum_{p_1,\dots,q_l=0}^{N-1} \sum_{m_1,\dots,n_l=-\infty}^{\infty} \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l})$$
$$\times \operatorname{cum}\big(Y_{i,k}; (i,k) \in \nu_1\big) \cdots \operatorname{cum}\big(Y_{i,k}; (i,k) \in \nu_l\big)$$
$$\times \exp\big(-\mathrm{i}\lambda_{k_1}(p_1 - q_1)\big) \cdots \exp\big(-\mathrm{i}\lambda_{k_l}(p_l - q_l)\big).$$

We now fix one indecomposable partition $\tilde{\nu}$ and assume without loss of generality that

$$\tilde{\nu} = \bigcup_{i=1}^{l-1} (Y_{i,1}, Y_{i+1,2}) \cup (Y_{l,1}, Y_{1,2}).$$

Because of $\operatorname{cum}(Z_i, Z_j) \neq 0$ for $i \neq j$, we obtain the following l equations:

$$q_1 = p_l + n_1 - m_l + t_{j_l} - t_{j_1}, \tag{A.11}$$
$$q_{i+1} = p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}} \qquad \text{for } i \in \{1, \dots, l-1\} \tag{A.12}$$

and therefore only l variables (namely p_i for $i \in \{1, \dots, l\}$) of the $2l$ variables $p_1, q_1, p_2, \dots, q_l$ are free to choose and must satisfy the following conditions:

$$0 \le p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}} \le N - 1 \qquad \text{for } i \in \{1, \dots, l-1\}, \tag{A.13}$$
$$0 \le p_l + n_1 - m_l + t_{j_l} - t_{j_1} \le N - 1. \tag{A.14}$$

Using the identities (A.11) and (A.12), we obtain that $V_T(\tilde{\nu})$ equals

$$\frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2} \phi(u_{j_1}, \lambda_{k_1}) \cdots \phi(u_{j_l}, \lambda_{k_l}) \frac{1}{(2\pi N)^l}$$
$$\times \sum_{p_1,p_2,\dots,p_l=0}^{N-1} \sum_{\substack{m_1,n_1,\dots,m_l,n_l=-\infty \\ (\text{A.13}),\ (\text{A.14})}}^{\infty} \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l}) \exp\big(-\mathrm{i}\lambda_{k_1}(p_1 - p_l)\big)$$
$$\times \prod_{i=1}^{l-1} \exp\big(-\mathrm{i}\lambda_{k_{i+1}}(p_{i+1} - p_i)\big)$$
$$\times \exp\big(-\mathrm{i}\lambda_{k_1}(m_l - n_1 + t_{j_1} - t_{j_l})\big)$$
$$\times \prod_{i=1}^{l-1} \exp\big(-\mathrm{i}\lambda_{k_{i+1}}(m_i - n_{i+1} + t_{j_{i+1}} - t_{j_i})\big).$$

We rename the m_i, n_i (m_i is replaced by n_i and n_i is replaced with m_{i-1} where we identify $l+1$ with 1 and 0 with l). Then (A.13) and (A.14) become

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \qquad \text{for } i \in \{1, \dots, l-1\}, \tag{A.15}$$

$$0 \le p_l + m_l - n_l + t_{j_l} - t_{j_1} \le N-1 \tag{A.16}$$

and after a factorisation in the arguments of the exponentials we obtain that $V_T(\tilde{\nu})$ is equal to

$$\begin{aligned}
&\frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi(u_{j_1}, \lambda_{k_1}) \cdots \phi(u_{j_l}, \lambda_{k_l}) \frac{1}{(2\pi N)^l} \\
&\quad \times \sum_{p_1,p_2,\dots,p_l=0}^{N-1} \sum_{\substack{m_1,n_1,\dots,m_l,n_l=-\infty \\ \text{(A.15), (A.16)}}}^{\infty} \psi_{m_1}(u_{j_2}) \cdots \psi_{n_l}(u_{j_l}) \\
&\qquad \times \prod_{i=1}^{l-1} \exp\bigl(-\mathrm{i}(\lambda_{k_i} - \lambda_{k_{i+1}}) p_i\bigr) \exp\bigl(-\mathrm{i}(\lambda_{k_l} - \lambda_{k_1}) p_l\bigr) \\
&\qquad \times \exp\bigl(-\mathrm{i}\lambda_{k_1}(n_l - m_l + t_{j_1} - t_{j_l})\bigr) \\
&\qquad \times \prod_{i=1}^{l-1} \exp\bigl(-\mathrm{i}\lambda_{k_{i+1}}(n_i - m_i + t_{j_{i+1}} - t_{j_i})\bigr).
\end{aligned}$$

We see that one can divide the sum with respect to p_i, m_i, n_i into a product of two sums, namely one sum with respect to all p_i, m_i, n_i with even i and the same sum with odd i. Analogously, we divide (A.15) and (A.16) into

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \qquad \text{for } i \in \{1, 3, 5, \dots, l-3, l-1\} \tag{A.17}$$

and

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \qquad \text{for } i \in \{2, 4, 6, \dots, l-4, l-2\}, \tag{A.18}$$

$$0 \le p_l + m_l - n_l + t_{j_l} - t_{j_1} \le N-1. \tag{A.19}$$

After applying the Cauchy–Schwarz inequality we obtain that $V_T(\tilde{\nu})$ is bounded by

$$\begin{aligned}
&\Biggl\{ \frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi(u_{j_1}, \lambda_{k_1})^2 \phi(u_{j_3}, \lambda_{k_3})^2 \cdots \phi(u_{j_{l-1}}, \lambda_{k_{l-1}})^2 \frac{1}{(2\pi N)^l} \\
&\quad \times \Biggl| \sum_{p_1=0}^{N-1} \exp\bigl(-\mathrm{i}(\lambda_{k_1} - \lambda_{k_2}) p_1\bigr) \sum_{p_3=0}^{N-1} \exp\bigl(-\mathrm{i}(\lambda_{k_3} - \lambda_{k_4}) p_3\bigr) \cdots
\end{aligned} \tag{A.20}$$

$$\times \sum_{p_{l-1}=0}^{N-1} \exp\bigl(-\mathrm{i}(\lambda_{k_{l-1}} - \lambda_{k_l}) p_{l-1}\bigr)$$

$$\times \sum_{\substack{m_1,n_1,m_3,n_3,\dots,m_{l-1},n_{l-1}=-\infty \\ (\mathrm{A.17})}}^{\infty} \psi_{m_1}(u_{j_2})\psi_{n_1}(u_{j_1})\psi_{m_3}(u_{j_4})\psi_{n_3}(u_{j_3})\cdots \times \psi_{m_{l-1}}(u_{j_l})\psi_{n_l}(u_{j_{l-1}})$$

$$\times \prod_{a\in\{1,3,\dots,l-1\}} \exp\bigl(-\mathrm{i}\lambda_{k_{a+1}}(n_a - m_a + t_{j_{a+1}} - t_{j_a})\bigr)\Bigr|^2\Bigr\}^{1/2}$$

$$\times \{\text{the same term with even } p_i, m_i, n_i\}^{1/2}.$$

We only consider the first term in (A.20), which is equal to

$$J_T := \frac{1}{2^l T^{l/2}} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2} \phi(u_{j_1},\lambda_{k_1})^2 \phi(u_{j_3},\lambda_{k_3})^2 \cdots \phi(u_{j_{l-1}},\lambda_{k_{l-1}})^2 \times \frac{1}{(2\pi N)^l} \bigl|K_T(u_1,\dots,u_l,\lambda_{k_1},\dots,\lambda_{k_l})\bigr|^2 \tag{A.21}$$

with $K_T(u_1,\dots,u_l,\lambda_{k_1},\dots,\lambda_{k_l})$ being defined implicitly. We have

$$\frac{1}{(2\pi N)^l} \sum_{k_2,k_4,\dots,k_l=-\lfloor (N-1)2\rfloor}^{N/2} \bigl|K_T(u_1,\dots,u_l,\lambda_{k_1},\dots,\lambda_{k_l})\bigr|^2$$

$$= \frac{1}{(2\pi N)^l} \sum_{p_1,p_3,\dots,p_{l-1}=0}^{N-1} \sum_{\tilde p_1,\tilde p_3,\dots,\tilde p_{l-1}=0}^{N-1} \sum_{\substack{m_1,n_1,m_3,n_3,\dots,m_{l-1},n_{l-1}=-\infty \\ (\mathrm{A.17})}}^{\infty} \sum_{\substack{\tilde m_1,\tilde n_1,\tilde m_3,\tilde n_3,\dots,\tilde m_{l-1},\tilde n_{l-1}=-\infty \\ (\tilde{\mathrm{A.17}})}}^{\infty}$$

$$\exp\bigl(-\mathrm{i}\lambda_{k_1}(p_1-\tilde p_1)\bigr)\exp\bigl(-\mathrm{i}\lambda_{k_3}(p_3-\tilde p_3)\bigr)\cdots\exp\bigl(-\mathrm{i}\lambda_{k_{l-1}}(p_{l-1}-\tilde p_{l-1})\bigr)$$

$$\times \psi_{m_1}(u_{j_2})\psi_{n_1}(u_{j_1})\cdots\psi_{m_{l-1}}(u_{j_l})\psi_{n_{l-1}}(u_{j_{l-1}})\psi_{\tilde m_1}(u_{j_2})\psi_{\tilde n_1}(u_{j_1})\cdots$$

$$\times \psi_{\tilde m_{l-1}}(u_{j_l})\psi_{\tilde n_{l-1}}(u_{j_{l-1}})$$

$$\times \sum_{k_2,k_4,\dots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2} \exp\bigl(-\mathrm{i}\lambda_{k_2}(\tilde p_1 - p_1 + n_1 - m_1 + \tilde m_1 - \tilde n_1)\bigr)$$

$$\times \exp\bigl(-\mathrm{i}\lambda_{k_4}(\tilde p_3 - p_3 + n_3 - m_3 + \tilde m_3 - \tilde n_3)\bigr)\cdots$$

$$\times \exp\bigl(-\mathrm{i}\lambda_{k_l}(\tilde p_{l-1} - p_{l-1} + n_{l-1} - m_{l-1} + \tilde m_{l-1} - \tilde n_{l-1})\bigr)$$

and because of the well-known identity

$$\frac{1}{N}\sum_{k=-\lfloor (N-1)/2\rfloor}^{N/2}\exp(-\mathrm{i}\lambda_k t)=\begin{cases}1, & t=lN \text{ with } l\in\mathbb{Z},\\ 0, & \text{else,}\end{cases}$$

it follows that for every i only one of the p_i and $\tilde{p}_i$ can be chosen freely if the m_i, n_i are fixed. Furthermore, we can show with the same arguments as in the proof of (A.4) that because of (A.17) and (2.3) we only have to consider the cases with $j_i = j_{i+1}$ for every odd i and that all other terms are of order $\mathrm{O}(1/N)$. This implies

$$\frac{1}{(2\pi N)^l}\sum_{k_2,k_4,\ldots,k_l=-\lfloor (N-1)/2\rfloor}^{N/2}\left|K_T(u_1,\ldots,u_l,\lambda_{k_1},\ldots,\lambda_{k_l})\right|^2\le\frac{1}{(2\pi)^l}\left(\sum_{m=-\infty}^{\infty}|\psi_m|\right)^{2l}$$

with $|\psi| := \sup_u |\psi(u)|$, and since we only need to sum over j_i with odd i in (A.21), it follows

$$J_T\le\frac{1}{T^{l/2}(4\pi)^l}\left(\sum_{m=-\infty}^{\infty}|\psi_m|\right)^{2l}\left(\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi(u_j,\lambda_k)^2\right)^{l/2}\bigl(1+\mathrm{O}(1/N)\bigr).$$

We obtain the same upper bound for the second factor in (A.20) and this implies

$$\begin{aligned}\mathrm{cum}_l\bigl(\sqrt{T}\hat{D}_T(\phi)\bigr)&\le\sum_{\nu}\frac{1}{(4\pi)^l(2\pi)^{l/2}}\left(\sum_{m=-\infty}^{\infty}|\psi_m|\right)^{2l}\left(\int_0^1\int_0^{\pi}\phi^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\right)^{l/2}\\&\quad\times\bigl(1+\mathrm{O}\bigl(N^2/T^2\bigr)+\mathrm{O}(1/N)\bigr)\\&\le(2l)!2^l\frac{1}{(4\pi)^l(2\pi)^{l/2}}\left(\sum_{m=-\infty}^{\infty}|\psi_m|\right)^{2l}\left(\int_0^1\int_0^{\pi}\phi^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\right)^{l/2}\\&\quad\times\bigl(1+\mathrm{O}\bigl(N^2/T^2\bigr)+\mathrm{O}(1/N)\bigr)\\&\le(2l)!\tilde{C}^l\rho_2(\phi)^l,\end{aligned}$$

where the last inequality follows because of $N/T\to 0$ and $1/N\to 0$ and since $(2l)!2^l$ is an upper bound for the number of indecomposable partitions of (A.10) (see Dahlhaus [6]). □

A.2. Proof of Theorem 3.3

A consequence of assumption (2.3) and $\int_0^1 f(u,\lambda)\,\mathrm{d}u>0$ for all $\lambda\in[-\pi,\pi]$ together with Lemma 2.1 of Kreiss, Paparoditis and Politis [19] is that

$$\sum_{j=1}^{\infty} j|a_j|<\infty \tag{A.22}$$

holds, and Lemma 2.3 in Kreiss, Paparoditis and Politis [19] implies that there exists a $p_0 \in \mathbb{N}$ such that for all $p \geq p_0$ the AR(p) process defined through (3.4) has an MA(∞) representation

$$Y_t^{\mathrm{AR}}(p) = \sum_{l=0}^{\infty} \psi_l^{\mathrm{AR}}(p) Z_{t-l}^{\mathrm{AR}}(p). \tag{A.23}$$

Furthermore, (3.10) together with Lemma 2.3 in Kreiss, Paparoditis and Politis [19] imply that there exist a $p_0' \in \mathbb{N}$, such that for all $p \geq p_0'$ the fitted AR(p) process has an MA(∞) representation

$$X_{t,T}^* = \sum_{l=0}^{\infty} \hat{\psi}_l^{\mathrm{AR}}(p,T) Z_{t-l}^*, \tag{A.24}$$

and we assume without loss of generality that T and $p(T)$ are sufficiently large to ensure the existence of such a representation.

Recall the proof of Theorem 2.1. In case the process is stationary, all the terms of order $\mathrm{O}(N^2/T^2)$ and $\mathrm{O}(1/T)$ vanish, as they are due to certain approximation errors which do not appear for $\psi_{t,T,l} = \psi_l(u) = \psi_l$. For a fixed p and T, the process of interest (A.24) is now indeed a stationary one and therefore the proof of Theorem 3.3 works in the same way as the previous one, if the remaining terms (which are the ones of order $\mathrm{O}_P(1/N)$) are a $\mathrm{o}_P(T^{-1/2})$ for the bootstrap process as well. Even more precisely, we only need the terms of order $\mathrm{O}_P(1/N)$ to be a $\mathrm{o}_P(T^{-1/2})$ in the calculation of the expectation, while it would suffice that they are a $\mathrm{o}_P(1)$ in the calculation of the higher order cumulants. A detailed look at the proof of Theorem 2.1 reveals that these terms are up to a constant of the form

$$\frac{(\sum_{m=0}^{\infty} |\psi_m|)^{q_1} (\sum_{l=0}^{\infty} l|\psi_l|)^{q_2}}{N}$$

with $q_1, q_2 \in \mathbb{N}$. For example, if the process is stationary we obtain from (A.5) an upper bound for $|\mathbb{E}(\hat{D}_T(u,\lambda))|$ via

$$C \frac{\sum_{m=0}^{\infty} |\psi_m| \sum_{l=0}^{\infty} l|\psi_l|}{N} = \mathrm{O}(1/N)$$

for some $C \in \mathbb{R}$, so an upper bound for the expectation of the bootstrap analogue $\hat{D}_T^*(u,\lambda)$ of $\hat{D}_T(u,\lambda)$ is given by

$$C \frac{\sum_{m=0}^{\infty} |\hat{\psi}_m^{\mathrm{AR}}(p,T)| \sum_{l=0}^{\infty} l|\hat{\psi}_l^{\mathrm{AR}}(p,T)|}{N}.$$

Therefore, it needs to be shown that

$$\sqrt{T} \frac{\sum_{m=0}^{\infty} |\hat{\psi}_m^{\mathrm{AR}}(p,T)| \sum_{l=0}^{\infty} l|\hat{\psi}_l^{\mathrm{AR}}(p,T)|}{N} = \mathrm{o}_P(1)$$

holds to obtain

$$\sqrt{T}\mathbb{E}\big(\hat{D}_T^*(u,\lambda)\big) = \mathrm{o}_P(1).$$

Because of (3.10), we can use the following bound from the proof of Theorem 3.1 in Berg, Paparoditis and Politis [3] for the difference between $\hat{\psi}_l^{\mathrm{AR}}(p,T)$ and $\psi_l^{\mathrm{AR}}(p)$ which is uniform in $p(T)$ and in $l \in \mathbb{N}$:

$$\big|\hat{\psi}_l^{\mathrm{AR}}(p,T) - \psi_l^{\mathrm{AR}}(p)\big| \le p(1+1/p)^{-l}\mathrm{O}_P(\sqrt{\log T/T}). \tag{A.25}$$

With (A.25), we obtain

$$\sum_{l=0}^{\infty}\big|\hat{\psi}_l^{\mathrm{AR}}(p,T) - \psi_l^{\mathrm{AR}}(p)\big| = \mathrm{O}_P\big(p_{\max}^2(T)\sqrt{\log T/T}\big)$$

and

$$\sum_{l=0}^{\infty} l\big|\hat{\psi}_l^{\mathrm{AR}}(p,T) - \psi_l^{\mathrm{AR}}(p)\big| = \mathrm{O}_P\big(p_{\max}^3(T)\sqrt{\log T/T}\big)$$

using properties of the geometric series, which yields

$$\sum_{l=0}^{\infty}\big|\hat{\psi}_l^{\mathrm{AR}}(p,T)\big| \le \mathrm{O}_P\big(p_{\max}^2(T)\sqrt{\log T/T}\big) + \sum_{l=0}^{\infty}\big|\psi_l^{\mathrm{AR}}(p)\big|$$

and

$$\sum_{l=0}^{\infty} l\big|\hat{\psi}_l^{\mathrm{AR}}(p,T)\big| \le \mathrm{O}_P\big(p_{\max}^3(T)\sqrt{\log T/T}\big) + \sum_{l=0}^{\infty} l\big|\psi_l^{\mathrm{AR}}(p)\big|.$$

Lemma 2.4 of Kreiss, Paparoditis and Politis [19] now implies that

$$\sum_{l=1}^{\infty}(1+l)\big|\psi_l^{\mathrm{AR}}(p) - \psi_l\big| \le \tilde{C}\sum_{l=p+1}^{\infty}(1+l)|a_l| \tag{A.26}$$

for another constant $\tilde{C} \in \mathbb{R}$, where the a_l are the coefficients of the AR(∞) representation in (3.1). Note that we implicitly assumed in (A.26) that the ψ_l are the coefficients of the Wold representation of the process Y_t defined in (3.1), since this particular bound only holds for this special MA representation. However, since the proof of Theorem 2.1 does not depend at all on the kind of MA representation, we can assume without loss of generality that the ψ_l are the coefficients of the Wold representation, and then (A.26) together with (2.3) and (A.22) yields

$$\sum_{l=0}^{\infty} l\big|\psi_l^{\mathrm{AR}}(p)\big| \le \bar{C}$$

for $\bar{C} \in \mathbb{R}$. Therefore, we obtain with (3.9)

$$\left(\sum_{m=0}^{\infty}\left|\hat{\psi}_m^{\mathrm{AR}}(p,T)\right|\right)^{p_1}\left(\sum_{l=0}^{\infty}l\left|\hat{\psi}_l^{\mathrm{AR}}(p,T)\right|\right)^{p_2} = \mathrm{O}_P(1)$$

for $p_1, p_2 \in \mathbb{N}$, which yields the assertion.

Acknowledgements

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors would like to thank two referees and an Associate Editor for their constructive comments on an earlier version of this manuscript. We are also grateful to Martina Stein who typed parts of this manuscript with considerable technical expertise.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. Budapest: Akadémiai Kiadó. MR0483125

[2] Beltrão, K.I. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Series Anal.* **8** 21–38. MR0868015

[3] Berg, A., Paparoditis, E. and Politis, D.N. (2010). A bootstrap test for time series linearity. *J. Statist. Plann. Inference* **140** 3841–3857. MR2674170

[4] Brillinger, D.R. (1981). *Time Series*: *Data Analysis and Theory*. New York: McGraw-Hill.

[5] Chiann, C. and Morettin, P.A. (1999). Estimation of time varying linear systems. *Stat. Inference Stoch. Process.* **2** 253–285. MR1919869

[6] Dahlhaus, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Process. Appl.* **30** 69–83. MR0968166

[7] Dahlhaus, R. (1996). On the Kullback–Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62** 139–168. MR1388767

[8] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916

[9] Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *J. Econometrics* **151** 101–112. MR2559818

[10] Dahlhaus, R., Neumann, M.H. and von Sachs, R. (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* **5** 873–906. MR1715443

[11] Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.* **34** 2790–2824. MR2329468

[12] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797

[13] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352

[14] Dette, H., Preuss, P. and Vetter, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768

[15] Dwivedi, Y. and Subba Rao, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91. MR2790673
[16] Hannan, E. and Kavalieris, L. (1986). Regression, autoregression models. *J. Time Series Anal.* **7** 27–49. MR0832351
[17] Kreiß, J.P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Fachbereich Mathematik, Univ. Hamburg.
[18] Kreiß, J.P. (1997). Asymptotical properties of residual bootstrap for autoregressions. Technical report, TU Braunschweig.
[19] Kreiss, J.P., Paparoditis, E. and Politis, D.N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MR2893863
[20] Neumann, M.H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25** 38–76. MR1429917
[21] Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary Gaussian long-memory processes. *Ann. Statist.* **38** 2958–2997. MR2722461
[22] Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. MR2597589
[23] Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Amer. Statist. Assoc.* **105** 839–851. MR2724865
[24] Pollard, D. (1984). *Convergence of Stochastic Processes. Springer Series in Statistics*. New York: Springer. MR0762984
[25] Priestley, M.B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 204–237.
[26] Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semi-parametric hypotheses for locally stationary processes. *Scand. J. Stat.* **36** 800–821. MR2573309
[27] Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist.* **36** 1879–1924. MR2435459
[28] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. New York: Springer. MR1385671
[29] von Sachs, R. and Neumann, M.H. (2000). A wavelet-based test for stationarity. *J. Time Series Anal.* **21** 597–613. MR1794489
[30] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist and Wiksell.
[31] Whittle, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.* **35** 48–60. MR0049539

Received September 2011 and revised March 2012