A test for stationarity based on empirical processes

Philip Preuß, Mathias Vetter, Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: philip.preuss@ruhr-uni-bochum.de
email: mathias.vetter@ruhr-uni-bochum.de
email: holger.dette@ruhr-uni-bochum.de

September 16, 2011

Abstract

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 the one hand the test can detect local alternatives converging to the null hypothesis at a rate $1/\sqrt{T}$ (where T denotes the sample size). On the other hand the method only requires the specification of one regularization parameter. The finite sample properties of the method are investigated by means of a simulation study and a comparison with two other tests is provided which have been proposed in the literature for testing stationarity.

AMS subject classification: 62M10, 62M15, 62G10
Keywords and phrases: spectral density, non stationary processes, goodness-of-fit tests, empirical spectral measure, integrated periodogram, locally stationary process, bootstrap

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 (1965), 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 (1996) and particular important examples are time varying ARMA models.

While many estimation techniques for locally stationary processes were developed [see Neumann and von Sachs (1997), Dahlhaus et al. (1999), Chang and Morettin (1999), Dahlhaus and Polonik (2006), Dahlhaus and Subba Rao (2006), Van Bellegem and von Sachs (2008) or Palma and Olea (2010) among others], semiparametric testing has found much less attention although its importance was pointed out by many authors. von Sachs and Neumann (2000) 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 (2009) and Paparoditis (2010) 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 (2010) 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 methods require the choice of at least one additional regularization parameter. It was pointed out in Sergides and Paparoditis (2009) that it is the choice of this particular tuning parameter that can influence the results of the statistical analysis substantially (the procedure proposed by these authors uses an additional smoothing bandwidth for the estimation of the local spectral density).

Recently Dette et al. (2011) 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 (2009) by calculating integrals of powers of the spectral density directly via Riemann sums of the periodogram. By this idea, Dette et al. (2011) avoided the integration of the smoothed periodogram [as it was done in Paparoditis (2009) or Paparoditis (2010)]. 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 et al. (2011) has attractive features it can only detect local alternatives converging to the null hypothesis at a rate $T^{-1/4}$ (here and throughout this paper T denotes the sample size). It is the purpose of the present paper to develop a test for stationarity in locally stationary processes which can on the one hand detect alternatives converging to the null hypothesis at the rate $T^{-1/2}$ and on the other hand requires only the specification of one regularization parameter. For this purpose we employ a Kolmogorov-Smirnov-type test statistic to estimate a measure of deviation from stationarity, which is defined by

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

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

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

and $f(u,\lambda)$ denotes the time-varying spectral density. Note that the quantity D is obviously zero if the process is stationary (i.e. $f(u,\lambda)$ is does not depend on u). The consideration of functionals of the form (1.1) for the construction of a test for stationarity is very natural and was already suggested by Dahlhaus (2009). In particular, Dahlhaus and Polonik (2009) proposed an estimator of this quantity which is based on the integrated pre-periodogram (with respect to the Lebesgue measure). 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 (2009) that the asymptotic properties of an estimator based on Riemann approximation are 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 powers 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 absolute 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 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 an Riemann approximation for the integrated pre-periodogram proposed by Dahlhaus (2009) (see 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 the pre-periodogram in the Riemann approximations. In Section 4 we investigate the finite sample properties by means of a simulation study. Although the use of the pre-periodogram does not require the specification of any regularization parameter, it is demonstrated that it yields substantially less power compared to the statistic based on the localised periodogram. Additionally, it is also shown that the latter method is extremely robust with respect to different choices of the window width, which is used for the calcualtion of the localised periodogram. Moreover we also provide a comparison with the test proposed in Dette et al. (2011) and show that their proposal is outperformed by the new method in most cases. Finally, for the sake of a transparent presentation of the results all technical details are deferred to an appendix in Section 5.

2 The test statistic

Following Dahlhaus and Polonik (2009), 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

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

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 behaviour 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

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

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

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

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

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

are satisfied, which are in general rather mild [see Dette et al. (2011) for more details]. 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.

Define

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

then the function

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

is well defined and called the time varying spectral density of $\{X_{t,T}\}_{t=1,\dots,T}$ [see Dahlhaus (1996)]. 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. We then define the local periodogram at time u by

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

[see Dahlhaus (1997)], where we have set $X_{j,T} = 0$, if $j \notin \{1, \dots, T\}$. This is the usual periodogram computed from the observations $X_{\lfloor uT \rfloor - N/2+1,T}, \dots, X_{\lfloor uT \rfloor + N/2,T}$. It can be shown that

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

and therefore the statistic $I_N^X(u,\lambda)$ is an asymptotically unbiased estimator for the spectral density if $N \to 0$ and $N = o(T)$. 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

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

where the points

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

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

$$\lambda_k := \frac{2\pi k}{N}, \quad k = 1, ..., \frac{N}{2}$$

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

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

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

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

for which the representation

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

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 $O(\cdot)$-terms to vanish asymptotically after multiplication with $\sqrt{T}$. Therefore we define an empirical spectral process by

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

and assume

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

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

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

corresponding to the stationary case and the alternative. The proof is complicated and therefore deferred to the Appendix. Throughout this paper the symbol $\Rightarrow$ denotes weak convergence in $[0,1]^2$.

Theorem 2.1 *If the assumptions* (2.2)–(2.5) *and* (2.8) *are satisfied, then as* $T \to \infty$ *we have*

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

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

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

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

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

which yields

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

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 $\sup_{(v,\omega)\in[0,1]^2} |G(v,\omega)|$. The asymptotic properties under the alternative will imply 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

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

and depends on the unknown spectral density f. In order to avoid the estimation of the integral of 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 [asymptotically unbiased, but again not consistent] estimator for the time-varying spectral density is given by

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

which is called the pre-periodogram [see Neumann and von Sachs (1997)]. Based on this statistic we define an alternative process by

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

where $\lambda_{k,T} = \frac{2\pi k}{T}$. The convergence of the finite dimensional distributions of the process $(H_T^1(v,\omega))_{(v,\omega)\in[0,1]^2}$ has already been shown in Dahlhaus (2009). Tightness can be shown using similar arguments as given in the Appendix for the proof of Theorem 2.1, which are not given here for the sake of brevity. As a consequence we obtain the following result.

Theorem 2.2 *If the assumptions* (2.2)–(2.5) *and* (2.8) *are satisfied, then as* $T \to \infty$ *we have*

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

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. Moreover, 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 (2009), 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

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

instead of $(H_T^1(v,\omega))_{(v,\omega)\in[0,1]^2}$ and proved its weak convergence. Note also that asymptotic tightness has neither been studied for an integrated nor for a summarized local periodogram in the literature so far. Moreover, many other asymptotic results are only shown for the integral of the local periodogram or pre-periodogram instead of the sum over the Fourier coefficients [see for example Dahlhaus (1997) or Paparoditis (2010)]. 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

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

because of the Riemann approximation error, this fact is in general not true, as the derivative $\frac{\partial I_N^X(u,\lambda)}{\partial\lambda}$ 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 the integrated local periodogram or pre-periodogram can not be directly transferred to corresponding Riemann approxiamtions. These difficulties were also explicitly pointed out in Example 2.7 of Dahlhaus (2009).

Remark 2.4 A careful inspection of the proofs in the Appendix shows that (2.10) also holds in the case where

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

if $g_T = o(1/\sqrt{T})$. Here k is an appropriate function such that (2.14) defines a time-varying spectral density. Moreover, if $g_T = \frac{1}{\sqrt{T}}$, 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

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

(note that $f(u,\lambda)$ is replaced by $\frac{1}{\sqrt{T}}k(u,\lambda)$ in the definition of $D_{N,M}$). 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). 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 a rate $T^{-1/2}$. In contrast, the proposal of Dette et al. (2011) is based on an L_2-distance between $f(u,\lambda)$ and $\int_0^1 f(v,\lambda)dv$ and is therefore only able to detect alternatives converging to the null hypothesis at a rate $T^{-1/4}$.

3 Bootstrapping the test statistic

To approximate the limiting distribution of $\sup_{(v,\omega)\in[0,1]^2}|G(v,\omega)|$, we employ an $AR(\infty)$-bootstrap approximation, which was introduced by Kreiß (1988). The bootstrap works by fitting an $AR(p)$-model

($p \in I\!N$) to the data $X_{1,T}, ..., X_{T,T}$, where the parameter $p = p(T)$ increases with the sample size T. To be precise we first calculate an estimator $(\hat{a}_{1,p}, ..., \hat{a}_{p,p})$ for

$$(a_{1,p}, ..., a_{p,p}) = \underset{b_{1,p},...,b_{p,p}}{\operatorname{argmin}} \mathbb{E}\Big(X_{t,T} - \sum_{j=1}^{p} b_{j,p} X_{t-j,T}\Big)^2 \tag{3.1}$$

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

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

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

$$\hat{\sigma}_p^2 := \frac{1}{T-p} \sum_{t=p+1}^{T} (\hat{z}_t - \overline{z}_T)^2, \tag{3.2}$$

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

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

[in other words $\hat{\sigma}_p^2$ 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 $X_{1,T}, ..., X_{T,T}$ are replaced by the bootstrap replicates $X^*_{1,T}, ..., X^*_{T,T}$. To assure that this procedure approximates the limiting distribution corresponding to the null hypothesis both under the null hypothesis and the alternative, we define the stationary process $X_t^{AR}(p)$ as the process which is defined through

$$X_t^{AR}(p) = \sum_{j=1}^{p} a_{j,p} X_{t-j}^{AR}(p) + Z_t^{AR}(p),$$

where $Z_t^{AR}(p)$ is a Gaussian white noise process with mean zero and variance

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

where X_t denotes the stationary process with spectral density $\int_0^1 f(u, \lambda) du$. We now impose the following technical conditions:

Assumption 3.1

(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*

$$\frac{p^3_{max}(T)\sqrt{\log(T)}}{\sqrt{T}} = O(1) \tag{3.3}$$

(ii) The stationary process X_t with strictly positive spectral density $\int_0^1 f(u,\lambda)du$ has an $AR(\infty)$-representation, i.e.

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + Z_t^{AR} \tag{3.4}$$

where $(Z_j^{AR})_{j\in\mathbb{Z}}$ denotes a Gaussian white noise process with variance $\sigma^2 > 0$, $\sum_{j=1}^{\infty} |a_j| < \infty$ and

$$1 - \sum_{j=1}^{\infty} a_j z^j \neq 0 \text{ for } |z| \leq 1.$$

(iii) The estimators for the AR parameters defined by (3.1) satisfy

$$\max_{1\leq j\leq p} |\hat{a}_{j,p} - a_{j,p}| = O(\sqrt{\log(T)/T}) \tag{3.5}$$

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

(iv) The estimate $\hat{\sigma}_p^2$ defined in (3.2) converges in probability to $\sigma^2 > 0$.

All assumptions are rather standard in the framework of an $AR(\infty)$-bootstrap [see for example Kreiß (1997) or Berg et al. (2010)] and it follows from Lemma 2.3 in Kreiß et al. (2011) that there exists a $p_0 \in I\!N$ such that for all $p \geq p_0$ the $AR(p)$-process defined through (3.1) has an $MA(\infty)$-representation

$$X_t^{AR}(p) = \sum_{l=0}^{\infty} \psi_l^{AR}(p) Z_{t-l}^{AR}(p). \tag{3.6}$$

Furthermore assumption (3.5) and Lemma 2.3 in Kreiß et al. (2011) imply that there exist a $p_0' \in I\!N$, such that for all $p \geq p_0'$ the fitted $AR(p)$-process has an $MA(\infty)$-representation

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

Because of (2.8) and (3.3), assumption (3.5) is for example satisfied for the least squares or the Yule-Walker estimators [see Hannan and Kavalieris (1986)]. These estimates have also the desired property that the fitted $AR(p)$-process has an $MA(\infty)$-representation for every p, if at least two observations are different which is typically the case. Note that (2.3) together with Lemma 2.1 of Kreiß et al. (2011) imply

$$\sum_{j=1}^{\infty} j|a_j| < \infty, \tag{3.7}$$

which will be used in the proof of the following theorem.

Theorem 3.2 *If the assumptions* (2.2)–(2.5), (2.8) *and Assumption 3.1 are satisfied, then as* $T \to \infty$ *we have conditionally on* $X_{1,T}, ..., X_{T,T}$

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

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

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

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 $i = 1, ..., B$ where $\hat{G}_{T,1}^*(v,\omega), ..., \hat{G}_{T,B}^*(v,\omega)$ are the B bootstrap replicates of $\hat{G}_T(v,\omega)$. We then reject the null hypothesis, whenever

$$\sqrt{T} \sup_{(v,\omega)\in[0,1]^2} |\hat{D}_T(v,\omega)| > (\hat{D}_T^*)_{T,\lfloor(1-\alpha)B\rfloor}, \tag{3.8}$$

where $(\hat{D}_T^*)_{T,1}, ..., (\hat{D}_T^*)_{T,B}$ denotes the order statistic of $\hat{D}_{T,1}^*, ..., \hat{D}_{T,B}^*$. This test has asymptotic level α because of Theorem 3.2 and is consistent, since conditionally on $X_{1,T}, ..., X_{T,T}$ each bootstrap statistic $\sup_{(v,\omega)\in[0,1]^2} |\hat{G}_T^*(v,\omega)|$ converges to a non generate 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 in $\hat{D}_T$ the localised periodogram 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.

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 [see Akaike (1973)], which is defined by

$$\hat{p} = \mathrm{argmin}_p \frac{1}{T} \sum_{k=1}^{\frac{T}{2}} \Big(\log(f_{\hat{\theta}(p)}(\lambda_{k,T}) + \frac{I_T^X(\lambda_{k,T})}{f_{\hat{\theta}(p)}(\lambda_{k,T})} \Big) + p/T$$

in the context of stationary processes [see Whittle (1951) or Whittle (1952)]. 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 (3.8) with respect to different choices of the parameter N. In particular it will be demonstrated in several examples that the test is very robust with respect to different choices of N.

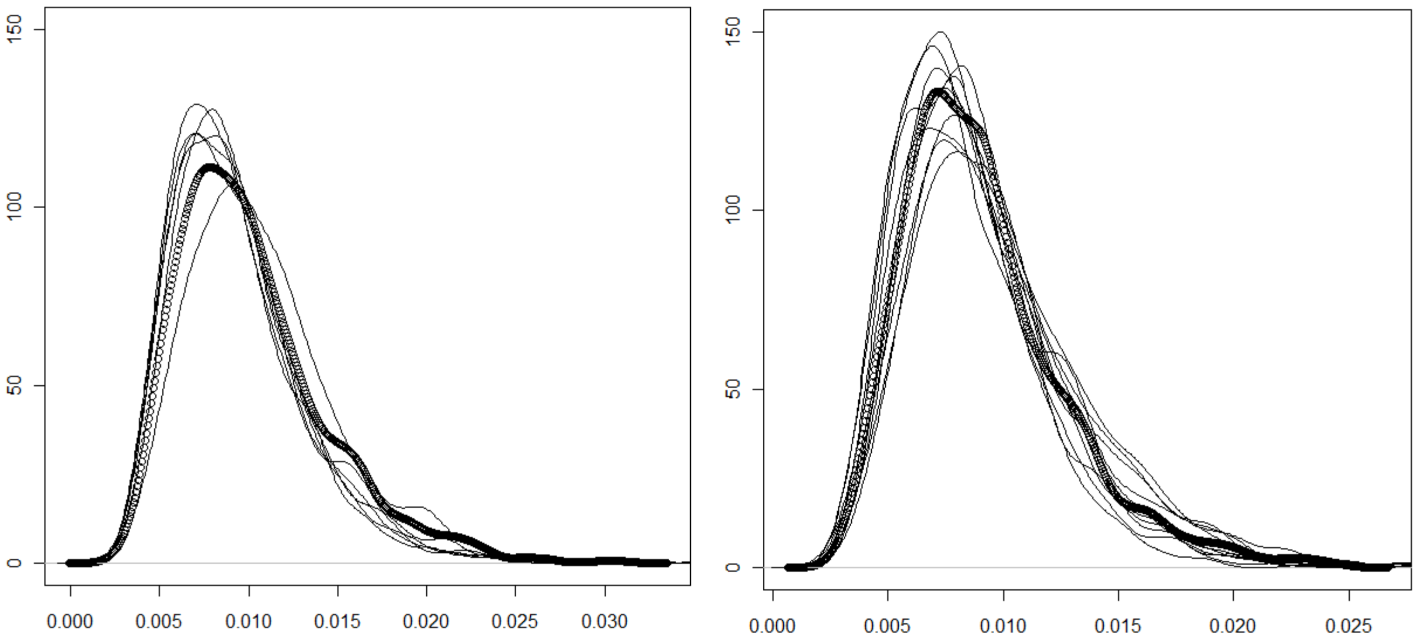

Figure 1: *Estimated densities of the distribution of the statistic* $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ *under the null hypothesis. The dotted line is the estimated exact density while the solid lines corresponds to the estimated densities of the bootstrap approximations. Left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

4.2 Bootstrap approximation

We now illustrate how 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 generated observations of the stationary $AR(1)$ model

$$X_{t,T} = 0.5X_{t-1,T} + Z_t \quad t = 1, ..., T$$

for $T = 128$ and calculated the bootstrap test statistic $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2}|\hat{D}_T(v,\omega)|$ both for $N = 16$ and $N = 8$. For both cases we generate 1000 replicates to estimate the exact distribution and chose randomly 10 series from the 1000 replications for which we calculate 1000 bootstrap approximations. Based on the 1000 bootstrap replications we estimate the density of the corresponding bootstrap approximation. The plots 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 (3.8) and the analogue based on the pre-periodogram. We also compare these methods with a test, which has recently been proposed by Dette et al. (2011). All reported results are based on 200 bootstrap replications and 1000 simulation

			$\phi = -0.5$		$\phi = 0$		$\phi = 0.5$	
T	N	M	5%	10%	5%	10%	5%	10%
64	8	8	0.025	0.06	0.035	0.086	0.05	0.099
128	16	8	0.031	0.077	0.042	0.081	0.034	0.092
128	8	16	0.03	0.076	0.038	0.083	0.055	0.102
256	32	8	0.04	0.086	0.051	0.106	0.053	0.111
256	16	16	0.038	0.089	0.044	0.085	0.045	0.08
256	8	32	0.036	0.083	0.051	0.098	0.05	0.102
512	64	8	0.054	0.103	0.052	0.084	0.042	0.09
512	32	16	0.046	0.083	0.044	0.09	0.049	0.092
512	16	32	0.038	0.079	0.056	0.098	0.052	0.099
512	8	64	0.05	0.102	0.047	0.101	0.051	0.112

Table 1: *Rejection probabilities of the test* (3.8) *under the null hypothesis. The data was generated according to model* (4.1).

runs under the null hypothesis while we used 500 simulation runs under the alternative. To study the approximation of the nominal level we simulate $AR(1)$ processes

(4.1) $$X_t = \phi X_{t-1} + Z_t,\ t \in \mathbb{Z}$$

and $MA(1)$ processes

(4.2) $$X_t = Z_t + \theta Z_{t-1},\ t \in \mathbb{Z}$$

for different values of the parameters ϕ and θ. The corresponding results are depicted in Table 1 and 2 and we observe a precise approximation of the nominal level for $\phi \in \{-0.5, 0, 0.5\}$ and $\theta = 0.5$ 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 $\theta = -0.5$ the nominal level is underestimated for smaller T but for $T = 512$ the approximation of the nominal level becomes much more precise and is robust with respect to different choices of the window width N if it is chosen according to the assumptions (2.8) (so basically N should be larger than M).
To study the power of the test (3.8) we simulated data from the following four models which corresponds

			$\theta = -0.5$		$\theta = 0.5$	
T	N	M	5%	10%	5%	10%
64	8	8	0.012	0.041	0.045	0.091
128	16	8	0.023	0.05	0.043	0.087
128	8	16	0.025	0.043	0.05	0.102
256	32	8	0.033	0.081	0.04	0.074
256	16	16	0.025	0.061	0.043	0.083
256	8	16	0.025	0.057	0.059	0.112
512	64	8	0.038	0.075	0.052	0.106
512	32	16	0.035	0.075	0.047	0.094
512	16	32	0.029	0.058	0.05	0.093
512	8	64	0.025	0.053	0.07	0.116

Table 2: *Rejection probabilities of the test* (3.8) *under the null hypothesis. The data was generated according to model* (4.2).

to the alternative of a non-stationary process

$$X_{t,T} = (1 + t/T)Z_t \tag{4.3}$$

$$X_{t,T} = -0.9\sqrt{\frac{t}{T}}X_{t-1,T} + Z_t \tag{4.4}$$

$$X_{t,T} = \begin{cases} 0.5X_{t-1} + Z_t & \text{if } 1 \le t \le \frac{T}{2}, \\ -0.5X_{t-1} + Z_t & \text{if } \frac{T}{2} + 1 \le t \le T. \end{cases} \tag{4.5}$$

$$X_{t,T} = \begin{cases} 0.5X_{t-1} + Z_t & \text{if } 1 \le t \le \frac{T}{2}, \\ 10Z_t & \text{if } \frac{T}{2} + 1 \le t \le \frac{T}{2} + \frac{T}{64}, \\ 0.5X_{t-1} + Z_t & \text{if } \frac{T}{2} + \frac{T}{64} + 1 \le t \le T. \end{cases} \tag{4.6}$$

The corresponding rejection probabilities are reported in Table 3 and we observe a reasonable behavior of the procedure in all considered cases. Under the alternative the bootstrap test (3.8) is also robust with respect to different choices of N. Note that even for the choice $M = 32$, $N = 8$, which clearly contradicts (2.8), the results are satisfying.

It might be of interest to compare these results with other tests for the hypothesis of stationarity which have been suggested in the literature. Because we are interested in procedures, which require as less as possible regularization we restrict ourselves to a comparison with two procedures. In Table 5 we present the rejection frequencies if we use the pre-periodogram [which was defined in (2.13)] instead of the local periodogram in our approach [see Theorem 2.2 and the discussion at the end of Section 2]. Recall that the use of the pre-periodogram does not require the specification of the value N, which specifies the

			(4.3)		(4.4)		(4.5)		(4.6)	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.286	0.444	0.186	0.328	0.168	0.27	0.368	0.456
128	16	8	0.686	0.772	0.396	0.546	0.308	0.466	0.656	0.732
128	8	16	0.624	0.758	0.382	0.578	0.410	0.548	0.648	0.744
256	32	8	0.958	0.974	0.672	0.814	0.742	0.912	0.908	0.938
256	16	16	0.942	0.978	0.698	0.814	0.640	0.806	0.926	0.950
256	8	32	0.944	0.970	0.760	0.868	0.672	0.808	0.910	0.930

Table 3: *Rejection probabilities of the test* (3.8) *for several alternatives.*

number of observations for the calculation of the local periodogram. This makes its use attractive for practitioners. However, the results of the simulation 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 under alternative (4.5). Based on these observations the Kolmogorov-Smirnov test based on the pre-periodogram can not be recommended.

In Table 4 we show the corresponding rejection probabilities for the test proposed in Dette et al. (2011), which is–to our best knowledge–the only available method with only one regularization parameter (namely N). All other methods require at least the specification of two parameters (usually the choice of a smoothing bandwidth and N). Moreover, in a detailed simulation study Dette et al. (2011) demonstrated that their method is superior to other proposals no matter how the additional smoothing bandwidths are chosen. These authors proposed to estimate the L_2-distance

$$\int_0^1 \int_0^\pi \Big(f(u,\lambda) - \int_0^1 f(v,\lambda)dv\Big)^2 d\lambda du$$

using sums of the (squared) periodogram. In order to provide a fair comparison between the two methods we also employed the AR(∞)-bootstrap to the corresponding test to generate critical values [note that without bootstrap the method of Dette et al. (2011) is much more sensitive with respect to different choices of N]. We observe that the new method also outperforms the test proposed by Dette et al. (2011) in the alternatives (4.3), (4.4) and (4.6). In most cases the differences are substantial. On the other hand for example (4.5) the procedure of Dette et al. (2011) has larger power if $T = 64$ and $T = 128$, but for $T = 256$ the new method performs better in this case as well. A comparison of the test proposed by Dette et al. (2011) with the Kolmogorov-Smirnov test based on the pre-periodogram shows no clear picture. For smaller sample sizes the test based on the estimation of the L_2 distance usually has larger power (except for model (4.3)), while the opposite can be observed for the sample size $T = 256$ (with an exception for the process (4.5), where the pre-periodogram test has nearly no power).

			(4.3)		(4.4)		(4.5)		(4.6)	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
64	8	8	0.116	0.196	0.188	0.232	0.250	0.344	0.352	0.456
128	16	8	0.106	0.16	0.256	0.33	0.370	0.552	0.520	0.610
128	8	16	0.168	0.268	0.220	0.286	0.432	0.566	0.528	0.606
256	32	8	0.378	0.498	0.282	0.412	0.746	0.922	0.772	0.844
256	16	16	0.208	0.368	0.276	0.41	0.618	0.794	0.834	0.898
256	8	32	0.224	0.338	0.300	0.418	0.582	0.744	0.890	0.932

Table 4: *Rejection probabilities of the test proposed Dette et al. (2011) for several alternatives (quantiles obtained by AR(∞)-bootstrap).*

	(4.3)		(4.4)		(4.5)		(4.6)	
T	5%	10%	5%	10%	5%	10%	5%	10%
64	0.188	0.340	0.080	0.202	0.022	0.056	0.288	0.438
128	0.552	0.702	0.216	0.392	0.036	0.116	0.680	0.752
256	0.938	0.968	0.580	0.734	0.080	0.176	0.912	0.938

Table 5: *Rejection probabilities of the test based on the pre-periodogram for several alternatives.*

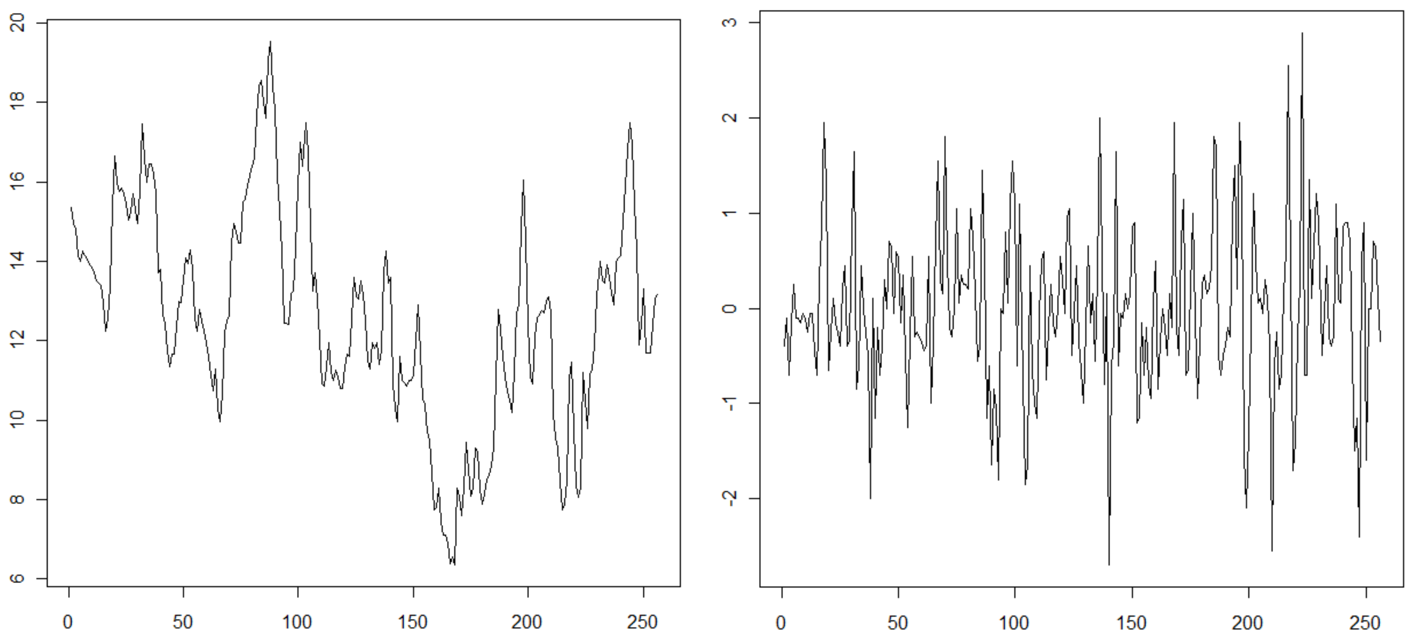

Figure 2: *Left panel: Weekly egg prices at a German agriculture market between April 1967 and March 1972. Right panel: First-order difference of the weekly egg prices.*

4.4 Data example

As an illustration we consider $T = 249$ 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 (2010) the first-order difference $\Delta_t = X_t - X_{t-1}$ of the observed time series are analyzed. Although in the literature several stationary models were proposed to fit this data [see Paparoditis (2010) for more details], the new test rejects the null hypothesis with the p-value 0.006 if we choose $N = 32$ or $N = 16$, and with the p-value 0.001 if we choose $N = 8$. These results are in line with the findings of Paparoditis (2010), and again the choice of N does not change the result much [note that the choice $N = 8$ contradicts to the assumption (2.8) and therefore one should use $N = 32$ or $N = 16$ which yields the same p-value].

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).

5 Appendix: Proofs

5.1 Proof of Theorem 2.1

To show weak convergence we will prove the following two claims [see van der Vaart and Wellner (1996), Theorem 1.5.4 and 1.5.7]:

(1) Convergence of the finite dimensional distributions

$$(\hat{G}_T(y_j))_{j=1,\ldots,K} \xrightarrow{D} (G(y_j))_{j=1,\ldots,K} \tag{5.1}$$

where $y_j = (v_j, \omega_j) \in [0,1]^2$ $(j = 1, \ldots, K)$ and $K \in \mathbb{N}$.

(2) Stochastic equicontinuity, i.e.

$$\forall \eta, \varepsilon > 0 \quad \exists \delta > 0 : \lim_{T\to\infty} P\Big(\sup_{y_1,y_2 \in [0,1]^2 : d_2(y_1,y_2)<\delta} |G_T(y_1) - G_T(y_2)| > \eta \Big) < \varepsilon. \tag{5.2}$$

Proof of (5.1)**:** The proof follows by similar arguments as given in the proof of Theorem 3.1 in Dette et al. (2011). For the sake of brevity and because we will use similar arguments in the proof of (5.2) we will sketch how the assertions

$$\mathbb{E}(\hat{G}_T(v,\omega)) \xrightarrow{T\to\infty} 0 \tag{5.3}$$

$$\mathrm{Cov}(\hat{G}_T(y_1), \hat{G}_T(y_2)) \xrightarrow{T\to\infty} \frac{1}{2\pi} \int_0^1 \int_0^{\pi \min(\omega_1,\omega_2)} (1_{[0,v_1]}(u) - v_1)(1_{[0,v_2]}(u) - v_2) f^2(u,\lambda) d\lambda du \tag{5.4}$$

with $y_j = (v_j, \omega_j)$ $(j = 1, 2)$ can be shown. Note that we have

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

with

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

for $u, \lambda \geq 0$ and

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

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 $u \geq 0, \lambda < 0$, and obtain

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

A Taylor expansion now yields that this is equal to

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

[for details see Dette et al. (2011)]. Since $\mathbb{E}(Z_iZ_j)=0$ for $i \neq j$ we obtain the equation $p = q+m-l$ which shows that the above expression equals

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

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

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

for some $C \in \mathbb{R}$ and the order follows from (2.3). Using (2.3) and (5.5) in the same way again, the first quantity above is equal to

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

and therefore we obtain

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

where the order of the Riemann approximation follows from the specific choice of the midpoints u_j. This together with (2.8) yields (5.3).
To prove (5.4) we use symmetry arguments and obtain

$$\begin{aligned}
&T\mathrm{cum}(\frac{1}{T}\sum_{j_1=1}^{M}\sum_{k_1=1}^{\frac{N}{2}}\phi_{v_1,\omega_1,M,N}(u_{j_1},\lambda_{k_1})I_N^X(u_{j_1},\lambda_{k_1}),\frac{1}{T}\sum_{j_2=1}^{M}\sum_{k_2=1}^{\frac{N}{2}}\phi_{v_2,\omega_2,M,N}(u_{j_2},\lambda_{k_2})I_N^X(u_{j_2},\lambda_{k_2}))\\
=&\frac{1}{4T}\frac{1}{(2\pi N)^2}\sum_{j_1,j_2=1}^{M}\sum_{k_1,k_2=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}}\phi_{v_1,\omega_1,M,N}(u_{j_1},\lambda_{k_1})\phi_{v_2,\omega_2,M,N}(u_{j_2},\lambda_{k_2})\\
&\times\sum_{p_1,p_2,q_1,q_2=0}^{N-1}\sum_{m_1,m_2,l_1,l_2=-\infty}^{\infty}\psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})\exp(-i\lambda_{k_1}(p_1-q_1))\exp(-i\lambda_{k_2}(p_2-q_2))\\
&\times\mathrm{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1}Z_{t_{j_1}-N/2+1+q_1-l_1},Z_{t_{j_2}-N/2+1+p_2-m_2}Z_{t_{j_2}-N/2+1+q_2-l_2})(1+O(N^2/T^2)+O(1/T))
\end{aligned}$$

in the same way as above. Because of

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

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

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

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 (5.5). 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 (2.3) and the well known identity

$$
\frac{1}{N}\sum_{k=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}}\exp(-i\lambda_k t)=\begin{cases}1, & t=lN \text{ with } l\in\mathbb{Z}\\ 0, & \text{else}\end{cases}\tag{5.6}
$$

that the corresponding terms are of order $O(1/N)$. The idea is that when all variables but q_1 and q_2 are fixed, then first there is for a given q_2 at most one choice for q_1 for which (5.6) becomes non-zero, and second the number of q_2 satisfying $0 \le q_2+m_1-l_2+t_{j_2}-t_{j_1} \le N-1$ is bounded by $|m_1-l_2|$ (if $j_1 \ne j_2$), so (2.3) can be applied.

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

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

Observing

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

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

$$
\Big| \sum_{\substack{q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1}}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})q_2) \Big| = \Big| \sum_{\substack{q_2=0 \\ q_2+m_1-l_2<0 \\ \text{or} \\ q_2+m_1-l_2>N-1}}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})q_2) \Big| \le |m_1 - l_2|,
$$

which implies

$$
(5.8) \qquad \Big| \frac{1}{(2\pi N)^2} \sum_{\substack{q_1,q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1 \\ 0 \le q_1+m_2-l_1 \le N-1}}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(q_2 - q_1)) \Big| \le |m_1 - l_2||m_2 - l_1|/(2\pi N)^2.
$$

By using (2.3) and (5.8) it can now be seen that all terms with $k_1 \neq k_2$ are of the order $O(1/N)$, and similar arguments as used in the calculation of the expectation yield that (5.7) equals

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

□

Proof of (5.2)**:** 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

$$
P_T := \Big\{ 0, \frac{1}{M}, \frac{2}{M}, ..., \frac{M-1}{M}, 1 \Big\} \times \Big\{ 0, \frac{2}{N}, \frac{4}{N}, ..., 1 - \frac{2}{N}, 1 \Big\}
$$

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

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

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 5.1 *There exists a constant $C \in I\!R$ such that for all $\phi \in \mathcal{F}_T^2$:*

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

Stochastic equicontinuity follows then by similar arguments as given in Dahlhaus (1988). To be precise, note that Theorem 5.1 implies the existence of a constant $C_1 \in I\!R$ such that for all $g, h \in \mathcal{F}_T$ and $\eta > 0$:

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

A straightforward modification of the chaining lemma in chapter VII.2 of Pollard (1984) yields that for a stochastic process $(Z(v))_{v \in V}$, whose index set V has a finite covering-integral

$$J(\delta) = \int_0^\delta \Big[\log\Big(\frac{48 N(u)^2}{u}\Big)\Big]^2 du \tag{5.9}$$

for all δ and which satisfies

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

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

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

In (5.9), $N(u)$ is the covering number which is defined as the smallest number $m \in I\!N$ for which there exist $z_1, ..., z_m \in V$ with $\min_i d(z, z_i) \le u$ for all $z \in V$. By using $y_i = (v_i, \omega_i)$ we obtain

$$P\Big(\sup_{y_1, y_2 \in P_T : d_2(y_1,y_2) < \delta} |\hat{G}_T(v_2, w_2) - \hat{G}_T(v_1, w_1)| > \eta\Big) \;\le\; P\Big(\sup_{f,g \in \mathcal{F}_T : \rho_2(f,g) < \varepsilon(\delta)} |\hat{G}_T(f) - \hat{G}_T(g)| > \eta\Big)$$

for $d_2(y_1, y_2) = \sqrt{(w_2 - w_1)^2 + (v_2 - v_1)^2}$ and a certain sequence $\varepsilon(\delta) \xrightarrow{\delta \to 0} 0$ by continuity. The right hand side of this inequality equals

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

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 $J(\varepsilon(\delta)) \stackrel{\delta\to 0}{\longrightarrow} 0$, we have $\eta > 26C_1 J(\delta)$ whenever δ is sufficiently small and obtain

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

which implies the stochastic equicontinuity.

Proof of Theorem 5.1: We show

$$|\mathrm{cum}_l(\sqrt{T}\hat{D}_T(\phi))| \leq (2l)!\tilde{C}^l \rho_2(\phi)^l \quad \forall l \in I\!N \tag{5.10}$$

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

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

for some C, and then it follows that (note that we consider only the case when k is even)

$$\mathbb{E}(|\hat{G}_T(\phi)|^k) = \Big|\sum_{\substack{\{P_1,\ldots,P_m\} \\ \text{Partition of} \\ \{1,\ldots,k\}}} \{\prod_{j=1}^{m} \mathrm{cum}_{|P_j|}(\hat{G}_T(\phi))\}\Big| \leq \rho_2(\phi)^k C^k \sum_{\substack{\{P_1,\ldots,P_m\} \\ \text{Partition of} \\ \{1,\ldots,k\}}} \prod_{j=1}^{m} (2|P_j|)! \leq (2k)! C^k 2^k \rho_2(\phi)^k,$$

[the last inequality follows from Dahlhaus (1988)] which yields the assertion.
In order to prove (5.10) we assume without loss of generality that l is even (the case for odd l is proved in the same way). The l-th cumulant of $\sqrt{T}\hat{D}_T(\phi)$ is given by

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

where both $O(\cdot)$-terms follow as in the proof of (5.3). 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 $i \in \{1,\ldots,l\}$. Theorem 2.3.2 in Brillinger (1981) yields

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

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

$$\begin{array}{cc} Y_{1,1} & Y_{1,2} \\ \vdots & \vdots \\ Y_{l,1} & Y_{l,2} \end{array} \tag{5.11}$$

and

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

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

$$q_1 = p_l + n_1 - m_l + t_{j_l} - t_{j_1} \tag{5.12}$$

$$q_{i+1} = p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}} \quad \text{for } i \in \{1, ..., l-1\} \tag{5.13}$$

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

$$0 \le p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}} \le N - 1 \quad \text{for } i \in \{1, ..., l-1\} \tag{5.14}$$

$$0 \le p_l + n_1 - m_l + t_{j_l} - t_{j_1} \le N - 1 \tag{5.15}$$

Using the identities (5.12) and (5.13), we obtain

$$\begin{aligned} V_T(\tilde{\nu}) =& \frac{1}{2^l T^{l/2}} \sum_{j_1,...,j_l=1}^{M} \sum_{k_1,...,k_l=-\lfloor \frac{N-1}{2} \rfloor}^{\frac{N}{2}} \phi(u_{j_1}, \lambda_{k_1}) \cdots \phi(u_{j_l}, \lambda_{k_l}) \frac{1}{(2\pi N)^l} \sum_{p_1,p_2,...,p_l=0}^{N-1} \sum_{\substack{m_1,n_1,...,m_l,n_l=-\infty \\ (5.14),(5.15)}}^{\infty} \\ & \times \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l}) \exp(-i\lambda_{k_1}(p_1 - p_l)) \prod_{i=1}^{l-1} \exp(-i\lambda_{k_{i+1}}(p_{i+1} - p_i)) \\ & \times \exp(-i\lambda_{k_1}(m_l - n_1 + t_{j_1} - t_{j_l})) \prod_{i=1}^{l-1} \exp(-i\lambda_{k_{i+1}}(m_i - n_{i+1} + t_{j_{i+1}} - t_{j_i})). \end{aligned}$$

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 (5.14) and (5.15) become

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \quad \text{for } i \in \{1, ..., l-1\} \tag{5.16}$$

$$0 \le p_l + m_l - n_l + t_{j_l} - t_{j_1} \le N-1 \tag{5.17}$$

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

$$\frac{1}{2^l T^{l/2}} \sum_{j_1,...,j_l=1}^{M} \sum_{k_1,...,k_l=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}} \phi(u_{j_1}, \lambda_{k_1}) \cdots \phi(u_{j_l}, \lambda_{k_l}) \frac{1}{(2\pi N)^l} \sum_{p_1,p_2,...,p_l=0}^{N-1} \sum_{\substack{m_1,n_1,...,m_l,n_l=-\infty \\ (5.16),(5.17)}}^{\infty} \psi_{m_1}(u_{j_2}) \cdots \psi_{n_l}(u_{j_l})$$

$$\prod_{i=1}^{l-1} \exp(-i(\lambda_{k_i} - \lambda_{k_{i+1}})p_i) \exp(-i(\lambda_{k_l} - \lambda_{k_1})p_l)$$

$$\exp(-i\lambda_{k_1}(n_l - m_l + t_{j_1} - t_{j_l})) \prod_{i=1}^{l-1} \exp(-i\lambda_{k_{i+1}}(n_i - m_i + t_{j_{i+1}} - t_{j_i}))$$

We see that one can divide the sum over the p_i, m_i, n_i into a product of two sums, namely one sum over all p_i, m_i, n_i with even i and the same sum with odd i. Analogously we divide (5.16) and (5.17) into

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \quad \text{for } i \in \{1, 3, 5, ..., l-3, l-1\} \tag{5.18}$$

and

$$0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N-1 \quad \text{for } i \in \{2, 4, 6, ..., l-4, l-2\} \tag{5.19}$$

$$0 \le p_l + m_l - n_l + t_{j_l} - t_{j_1} \le N-1. \tag{5.20}$$

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

$$\begin{aligned} &\Big\{ \frac{1}{2^l T^{l/2}} \sum_{j_1,...,j_l=1}^{M} \sum_{k_1,...,k_l=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}} \phi(u_{j_1}, \lambda_{k_1})^2 \phi(u_{j_3}, \lambda_{k_3})^2 \cdots \phi(u_{j_{l-1}}, \lambda_{k_{l-1}})^2 \frac{1}{(2\pi N)^l} \\ &\Big| \sum_{p_1=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})p_1) \sum_{p_3=0}^{N-1} \exp(-i(\lambda_{k_3} - \lambda_{k_4})p_3) \cdots \sum_{p_{l-1}=0}^{N-1} \exp(-i(\lambda_{k_{l-1}} - \lambda_{k_l})p_{l-1}) \\ &\qquad \sum_{\substack{m_1,n_1,m_3,n_3,...,m_{l-1},n_{l-1}=-\infty \\ (5.18)}}^{\infty} \psi_{m_1}(u_{j_2})\psi_{n_1}(u_{j_1})\psi_{m_3}(u_{j_4})\psi_{n_3}(u_{j_3}) \cdots \psi_{m_{l-1}}(u_{j_l})\psi_{n_l}(u_{j_{l-1}}) \\ &\qquad \prod_{a \in \{1,3,...,l-1\}} \exp(-i\lambda_{k_{a+1}}(n_a - m_a + t_{j_{a+1}} - t_{j_a})) \Big|^2 \Big\}^{1/2} \\ &\times \Big\{ \text{the same term with even } p_i, m_i, n_i \Big\}^{1/2} \end{aligned} \tag{5.21}$$

We only consider the first term in (5.21), which is equal to

(5.22)

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

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

$$\begin{aligned}
&\frac{1}{(2\pi N)^l} \sum_{k_2,k_4,...,k_l=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}} |K_T(u_1,...,u_l,\lambda_{k_1},...,\lambda_{k_l})|^2 \\
=&\frac{1}{(2\pi N)^l} \sum_{p_1,p_3,...,p_{l-1}=0}^{N-1} \sum_{\tilde{p}_1,\tilde{p}_3,...,\tilde{p}_{l-1}=0}^{N-1} \sum_{\substack{m_1,n_1,m_3,n_3,...,m_{l-1},n_{l-1}=-\infty \\ (5.18)}}^{\infty} \sum_{\substack{\tilde{m}_1,\tilde{n}_1,\tilde{m}_3,\tilde{n}_3,...,\tilde{m}_{l-1},\tilde{n}_{l-1}=-\infty \\ (5.\tilde{1}8)}}^{\infty} \\
&\exp(-i\lambda_{k_1}(p_1-\tilde{p}_1)) \exp(-i\lambda_{k_3}(p_3-\tilde{p}_3)) \cdots \exp(-i\lambda_{k_{l-1}}(p_{l-1}-\tilde{p}_{l-1})) \\
&\psi_{m_1}(u_{j_2})\psi_{n_1}(u_{j_1}) \cdots \psi_{m_{l-1}}(u_{j_l})\psi_{n_{l-1}}(u_{j_{l-1}})\psi_{\tilde{m}_1}(u_{j_2})\psi_{\tilde{n}_1}(u_{j_1}) \cdots \psi_{\tilde{m}_{l-1}}(u_{j_l})\psi_{\tilde{n}_{l-1}}(u_{j_{l-1}}) \\
&\sum_{k_2,k_4,...,k_l=-\lfloor\frac{N-1}{2}\rfloor}^{\frac{N}{2}} \exp(-i\lambda_{k_2}(\tilde{p}_1-p_1+n_1-m_1+\tilde{m}_1-\tilde{n}_1)) \exp(-i\lambda_{k_4}(\tilde{p}_3-p_3+n_3-m_3+\tilde{m}_3-\tilde{n}_3)) \\
&\cdots \exp(-i\lambda_{k_l}(\tilde{p}_{l-1}-p_{l-1}+n_{l-1}-m_{l-1}+\tilde{m}_{l-1}-\tilde{n}_{l-1}))
\end{aligned}$$

and because of (5.6) 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 (5.4) that because of (5.18) 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 $O(1/N)$. This implies

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

with $|\psi| := \sup_u |\psi(u)|$, and since we only need to sum over j_i with odd i in (5.22), it follows

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

We obtain the same upper bound for the second factor in (5.21) and this implies

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

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 (5.11) [see Dahlhaus (1988)]. □

5.2 Proof of Theorem 3.2

Let

$$X^*_{t,T} = \sum_{l=0}^{\infty}\hat{\psi}_l^{AR}(p)Z^*_{t-l} \qquad (5.23)$$

be the $MA(\infty)$ representation of the fitted $AR(p)$-model [its existence was shown in the discussion after Assumption 3.1 whenever T and thus $p(T)$ is sufficiently large]. If the process is stationary [i.e. $\psi_{t,T,l} = \psi_l(u) = \psi_l$], all the terms of order $O(N^2/T^2)$ and $O(1/T)$ vanish in the proof of Theorem 2.1. For a fixed p and T, the process (5.23) is stationary, and therefore the proof of Theorem 3.2 works in the same way as the previous one, if the (now random) terms of order $O_P(1/N)$ are a $o_P(T^{-1/2})$ for the bootstrap process as well [in fact we only need that the terms of order $O_P(1/N)$ are of order $o_P(T^{-1/2})$ in the calculation of the expectation while it would suffice that they are a $o_P(1)$ in the calculation of higher order cumulants]. Note that these terms in the proof of Theorem 2.1 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 we obtain from (5.5) [if the process is stationary] an upper bound for $|\mathbb{E}(\hat{D}_T(u,\lambda))|$ [where $\hat{D}_T(u,\lambda)$ was defined in (2.6)] via

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

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

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

Therefore it needs to be shown that

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

to obtain

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

Because of (3.5) we can use the following bound from the proof of Theorem 3.1. in Berg et al. (2010) for the difference between $\hat{\psi}_l^{AR}(p)$ and $\psi_l^{AR}(p)$ (where $\psi_l^{AR}(p)$ was defined in (3.6)) which is uniform in $p(T)$ and uniform in $l \in I\!N$:

$$|\hat{\psi}_l^{AR}(p) - \psi_l^{AR}(p)| \leq p(1+1/p)^{-l}O_P(\sqrt{\log T/T}) \tag{5.24}$$

With (5.24) we obtain

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

and

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

using properties of the geometric series, which yields

$$\sum_{l=0}^{\infty}|\hat{\psi}_l^{AR}(p)| \leq O_P(p_{max}^2(T)\sqrt{\log T/T}) + \sum_{l=0}^{\infty}|\psi_l^{AR}(p)|$$

and

$$\sum_{l=0}^{\infty}l|\hat{\psi}_l^{AR}(p)| \leq O_P(p_{max}^3(T)\sqrt{\log T/T}) + \sum_{l=0}^{\infty}l|\psi_l^{AR}(p)|.$$

Lemma 2.4 of Kreiß et al. (2011) now implies that

$$\sum_{l=1}^{\infty}(1+l)|\psi_l^{AR}(p) - \psi_l| \leq \tilde{C}\sum_{l=p+1}^{\infty}(1+l)|a_l| \tag{5.25}$$

for another constant $\tilde{C} \in I\!R$, where the a_l are the coefficients of the $AR(\infty)$-representation [see (3.4)]. Note that in (5.25) we implicitly assumed that the ψ_l are the coefficents of the Wold-representation of the process X_t defined in (3.4), since this is bound is only true for this special MA-representation. However, since the proof of Theorem 2.1 does not depend on the special type of MA-representation, we can assume without loss of generality that the ψ_l are the coefficents of the Wold-representation and then (5.25) together with (2.3) and (3.7) yields

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

for $\bar{C} \in I\!R$. Therefore we obtain with (3.3)

$$\Big(\sum_{m=0}^{\infty} |\hat{\psi}_m^{AR}(p)|\Big)^{p_1} \Big(\sum_{l=0}^{\infty} l|\hat{\psi}_l^{AR}(p)|\Big)^{p_2} = O_P(1)$$

for $p_1, p_2 \in I\!N$, which yields the assertion.

□

References

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. Budapest, Akademia Kiado, 267-281.

Berg, A., Paparoditis, E., and Politis, D. N. (2010). A bootstrap test for time series linearity. *Journal of Statistical Planning and Inference*, 140:3841–3857.

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

Chang, C. and Morettin, P. (1999). estimation of time-varying linear systems. *Statistical Inference for Stochastic Processes.*, 2:253–285.

Dahlhaus, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Process and their Applications*, 30:69–83.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process and their Applications*, 62:139–168.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics*, 151:101–112.

Dahlhaus, R., Neumann, M., and von Sachs, R. (1999). Nonlinear wavelet estimation of time varying autoregressive processes. *Bernoulli*, 5:873–906.

Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *Annals of Statistics*, 34(6):2790–2824.

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

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying arch processes. *Annals of Statistics.*, 34(3):1075–1114.

Dette, H., Preuß, P., and Vetter, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *to appear in Journal of the American Statistical Association.*

Dwivedi, Y. and Subba Rao, S. (2010). A test for second order stationarity of a time series based on the discrete fourier transform. *Journal of Time Series Analysis.*, 32(1):68–91.

Hannan, E. and Kavalieris, L. (1986). Some results in time series analysis. *Journal of Time Series Analysis*, 7.

Kreiß, J.-P. (1988). *Asymptotic statistical inference for a class of stochastic processes.* Habilitationsschrift, Fachbereich Mathematik, Universitï¿½t Hamburg.

Kreiß, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregressions. Technical report, TU Braunschweig.

Kreiß, J.-P., Paparoditis, E., and Politis, D. N. (2011). On the range of the validity of the autoregressive sieve bootstrap. *to appear in Annals of statistics.*

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and applications to adaptive estimation of evolutionary spectra. *Annals of Statistics*, 25:38–76.

Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary Gaussian long-memory processes. *Annals of Statistics.*, 38(5):2958–2997.

Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli*, 15:1190–1221.

Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association*, 105(490):839–851.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Ser. B*, 62:204–237.

Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semiparametric hypotheses for locally stationary processe. *Scandinavian Journal of Statistics*, 36:800–821.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics*, 36(4):1879–1924.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, Berlin.

von Sachs, R. and Neumann, M. H. (2000). A wavelet-based test for stationarity. *Journal of Time Series Analysis*, 21:597–613.

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. HUppsala: Almqvist and Wiksell.

Whittle, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.*, 35:48–60.