

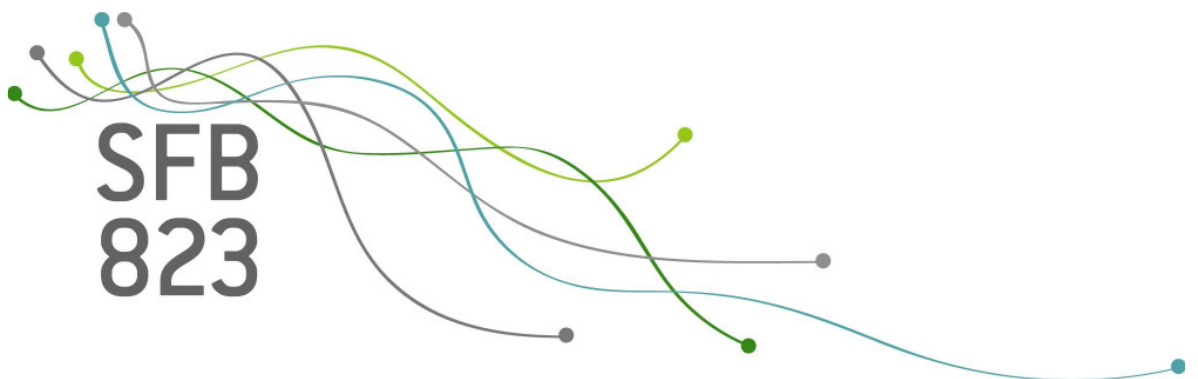
SFB
823

A measure of stationarity in locally stationary processes with applications to testing

Holger Dette, Philip Preuß, Mathias Vetter

Nr. 32/2010

Discussion Paper



A measure of stationarity in locally stationary processes with applications to testing

Holger Dette, Philip Preuß, Mathias Vetter

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

email: holger.dette@ruhr-uni-bochum.de

FAX: +49 2 34 32 14 559

July 28, 2010

Abstract

In this paper we investigate the problem of measuring deviations from stationarity in locally stationary time series. Our approach is based on a direct estimate of the L^2 -distance between the spectral density of the locally stationary process and its best approximation by a spectral density of a stationary process. An explicit expression of the minimal distance is derived, which depends only on integrals of the spectral density of the stationary process and its square. These integrals can be estimated directly without estimating the spectral density, and as a consequence, the estimation of the measure of stationarity does not require the specification of smoothing parameters. We show weak convergence of an appropriately standardized version of the statistic to a standard normal distribution. The results are used to construct confidence intervals for the measure of stationarity and to develop a new test for the hypothesis of stationarity which does not require regularization. Finally, we investigate the finite sample properties of the resulting confidence intervals and tests by means of a small simulation study and illustrate the methodology in three data examples.

AMS subject classification: 62M10, 62M15, 62G10

Keywords and phrases: spectral density, non stationary processes, goodness-of-fit tests, L^2 -distance, integrated periodogram, locally stationary process

1 Introduction

Locally stationary time series models have found considerable interest in the recent literature, because in many applications time series change their dependence characteristics as time evolves. These phenomena cannot be adequately described by the assumption of weak stationarity, and locally stationary processes provide an interesting class of models with more flexibility. These processes have been introduced as a more realistic theoretical framework for the analysis of time series which allows for the second-order characteristics of the underlying stochastic process, and, more specifically, for its auto covariance structure to vary with time. Out of the large literature we mention the early work on this subject of Priestley (1965), who considered oscillating processes. Neumann and von Sachs (1997) and Nason et al. (2000) discussed the estimation of evolutionary spectra by wavelet methods. Dahlhaus (1997) gave a definition of locally stationary processes on the basis of a time varying spectral representation and established the asymptotic theory for statistical inference in such cases [see also Dahlhaus (2000)]. Some applications of locally stationary processes to speech signals and earthquake data can be found in Adak (1998), while Sakiyama and Taniguchi (2004) discussed the problem of discriminant analysis for locally stationary processes. More recent work in this field can be found in Dahlhaus and Polonik (2006, 2009) and Dahlhaus (2009) who discussed quasi maximum likelihood estimation, empirical process theory and its application to statistical inference in locally stationary processes.

Several models for locally stationary processes have been proposed in the literature, including time varying $AR(p)$ models and time varying $ARMA(p, q)$ models. In contrast to the “classical inference” mentioned in the previous paragraph, the problem of testing semiparametric hypotheses (such as time varying autoregressive structure or stationarity) for a time varying spectral density has found much less attention in the literature. Sergides and Paparoditis (2009) investigated semiparametric hypotheses and proposed a bootstrap test in this context. Several authors have pointed out the importance of validating stationarity in locally stationary processes, such that the statistician is able to decide at an early stage whether an observed time series can be considered as covariance stationary or not. Sakiyama and Taniguchi (2003) considered the problem of testing stationarity versus local stationarity in a parametric locally stationary model, while Lee et al. (2003) investigated the constancy over time of a finite number of autocovariances. von Sachs and Neumann (2000) proposed a multiple testing procedure based on empirical wavelet coefficients estimated using localized versions of the periodogram, while Paparoditis (2010) used L_2 -distances between the local sample spectral density and an overall spectral density estimator [see also Paparoditis (2009)]. A common feature in many of these methods is the fact that the statistical inference depends on the choice of a regularization parameter. For example, Paparoditis (2009) and Paparoditis (2010) compare nonparametric estimators of the spectral density of the stationary and locally stationary process, and as a consequence, the resulting statistical analysis depends sensitively on the choice of a smoothing parameter which is required for the density estimation. An alternative approach in this context is the application of the empirical spectral measure for inference in locally stationary time series [see Dahlhaus and Polonik (2009)]. In particular Dahlhaus (2009) proposed a test for stationarity by comparing estimates of the integrated time frequency spectral density under the null hypothesis of stationarity and the alternative of local stationarity. This approach avoids

smoothing and under the null hypothesis the corresponding empirical process converges weakly to a Gaussian process. However, as pointed out in Example 2.7 of Dahlhaus (2009), the calculation of the limiting distribution of a corresponding Kolmogorov-Smirnov statistic is an unsolved task, because the limiting process depends in a complicated way on certain features of the data generating process.

The present paper is devoted to an extremely simple alternative method for measuring deviations from stationarity in locally stationary processes. We propose a measure for stationarity by the best L^2 -approximation of the spectral density of the underlying process by the spectral density of a stationary process. More precisely, we consider the minimal distance

$$(1.1) \quad D^2 = \min_g \int_{-\pi}^{\pi} \int_0^1 (f(u, \lambda) - g(\lambda))^2 du d\lambda,$$

where $f(u, \lambda)$ denotes the spectral density of the locally stationary process ($u \in [0, 1]$, $\lambda \in [-\pi, \pi]$) and the minimum is calculated over the set of all spectral densities g corresponding to stationary processes. Note that $D^2 = 0$ if and only if there exists a function $f : [-\pi, \pi] \rightarrow \mathbb{C}$, such that the hypothesis

$$(1.2) \quad H_0 : f(u, \lambda) = f(\lambda) \text{ a.e. on } [0, 1] \times [-\pi, \pi]$$

is satisfied, i.e. the given locally stationary process is in fact stationary. On the other hand, if the process is not stationary, D^2 could be considered as a measure for the deviation of the locally stationary process from stationarity. It will be shown in Section 2 that the minimal L^2 -distance defined in (1.1) can be determined explicitly and depends only on integrals of the functions $f(u, \lambda)$ and $f^2(u, \lambda)$ calculated over the full time and frequency domain, which can easily be estimated from the data by appropriate summations over local periodograms. As a consequence, we obtain an empirical measure of stationarity which avoids the problem of smoothing the local periodogram. Moreover, it can be shown that the limiting distribution of this estimate [after an appropriate standardization] is normal, where the corresponding asymptotic variance can easily be estimated from the data.

The remaining part of the paper is organized as follows. In Section 2 we introduce the necessary notation, the basic assumptions and explain the main principle of our approach. The asymptotic theory is derived in Section 3, while the finite sample properties of the estimate for the quantity D^2 are studied in Section 4. In particular, we investigate the coverage probability and the power of the constructed confidence intervals and tests. We also illustrate the methodology by re-analyzing several data examples, which have been recently discussed in the literature. Finally, some more technical details required in the asymptotic analysis are deferred to an appendix in Section 5.

2 Measuring stationarity

Let $\{X_{t,T}\}_{t=1,\dots,T}$ ($T \in \mathbb{N}$) denote a sequence of stochastic processes with the representation

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

where the random variables Z_t are assumed to be independent and identically normal distributed, having mean zero and variance σ^2 . The assumption of Gaussianity of the errors is imposed to simplify technical arguments [see Remark 3.5]. The quantities $\psi_{t,T,l}$ denote constants which satisfy

$$(2.2) \quad \sum_{l=-\infty}^{\infty} |\psi_{t,T,l}| < \infty$$

and are chosen in such a way that there exist twice continuously differentiable functions $\psi_l : [0, 1] \rightarrow \mathbb{R}$ with

$$(2.3) \quad \sum_{l=-\infty}^{\infty} \sup_t |\psi_{t,T,l} - \psi_l(t/T)| = O(1/T).$$

Throughout this paper we assume that the conditions

$$(2.4) \quad \sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} |\psi_l(u)| |l|^2 < \infty,$$

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

$$(2.6) \quad \sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} |\psi_l''(u)| < \infty$$

are satisfied. The time-varying spectral density of the locally stationary process $\{X_{t,T}\}$ is defined in terms of the auxiliary functions ψ_l , that is

$$(2.7) \quad f(u, \lambda) = \frac{\sigma^2}{2\pi} |\psi(u, \exp(-i\lambda))|^2,$$

where the function ψ is given by

$$(2.8) \quad \psi(u, \exp(-i\lambda)) := \sum_{l=-\infty}^{\infty} \psi_l(u) \exp(-i\lambda l).$$

Existence of the time-varying spectral density function follows from condition (2.4), and it is shown in Dahlhaus (1996) that the time varying spectral density f is unique under the assumptions stated in (2.4)–(2.6).

The following Lemma provides an explicit expression for the minimal distance between the local stationary density $f(u, \lambda)$ and the class of all spectral densities corresponding to stationary processes.

Lemma 2.1 *The minimal distance defined in (1.1) is given by*

$$(2.9) \quad D^2 = \int_{-\pi}^{\pi} \int_0^1 f^2(u, \lambda) du d\lambda - \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \right)^2 d\lambda$$

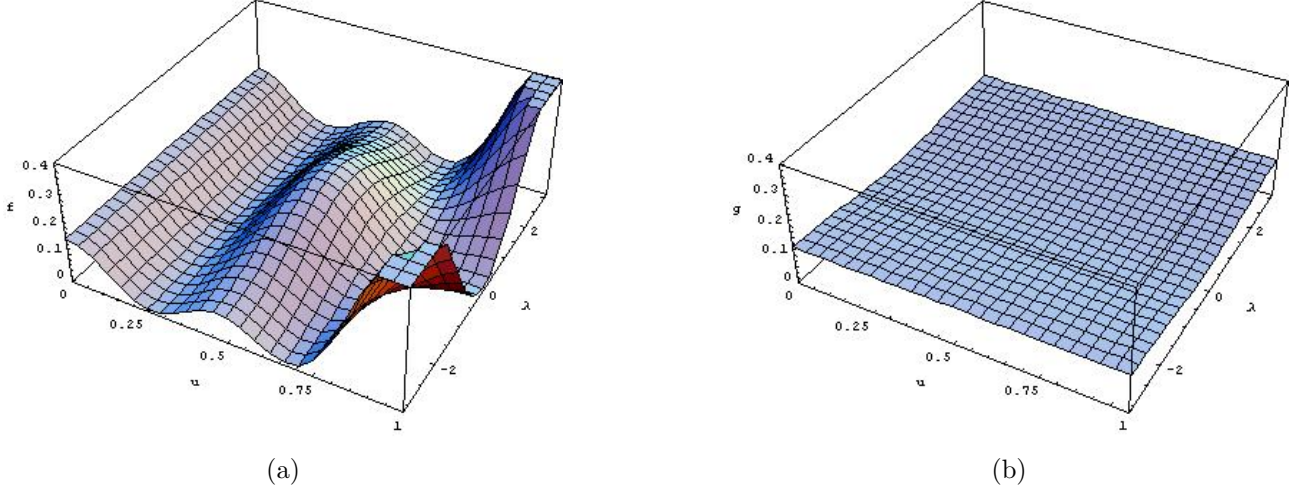


Figure 1: (a) Plot of the local spectral density $f(u, \lambda)$ in (2.11). (b) Plot of the best approximation of f by the spectral density $g(u, \lambda) = g^*(\lambda) = \int_0^1 f(u, \lambda) du$.

Proof. Let $g^*(\lambda) = \int_0^1 f(u, \lambda) du$, then we obtain

$$\begin{aligned}
 \int_{-\pi}^{\pi} \int_0^1 (f(u, \lambda) - g(\lambda))^2 du d\lambda &= \int_{-\pi}^{\pi} \int_0^1 (f(u, \lambda) - g^*(\lambda))^2 du d\lambda + \int_{-\pi}^{\pi} (g(\lambda) - g^*(\lambda))^2 d\lambda \\
 &\geq \int_{-\pi}^{\pi} \int_0^1 (f(u, \lambda) - g^*(\lambda))^2 d\lambda \\
 &= \int_{-\pi}^{\pi} \int_0^1 f^2(u, \lambda) du d\lambda - \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \right)^2 d\lambda,
 \end{aligned}$$

where there is equality if and only if $g = g^*$. □

Example 2.1 Consider the tvMA(2) process

$$(2.10) \quad X_{t,T} = \cos(2\pi t/T) Z_t - (t/T)^2 Z_{t-1},$$

where $\sigma^2 = 1$. We obtain by a straightforward calculation

$$(2.11) \quad f(u, \lambda) = \frac{1}{2\pi} \left\{ \cos(2\pi u)^2 - 2u^2 \cos(2\pi u) \cos(\lambda) + u^4 \right\},$$

and the best approximation via a stationary spectral density is given by

$$(2.12) \quad g^*(\lambda) = \int_0^1 f(u, \lambda) du = \frac{7}{20\pi} - \frac{1}{2\pi^3} \cos(\lambda).$$

Plots of the functions $f(u, \lambda)$ and $g^*(\lambda)$ are shown in Figure 1.

Observing the representation of the quantity D^2 in Lemma 2.1, an estimate for it can easily be constructed by estimating the integrals

$$(2.13) \quad F_1 = \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} f^2(u, \lambda) d\lambda du,$$

$$(2.14) \quad F_2 = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \right)^2 d\lambda.$$

For this purpose we will use local periodograms and 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. Then we define the local periodogram by

$$(2.15) \quad I_N^X(u, \lambda) := |J_N^X(u, \lambda)|^2,$$

where

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

[see Dahlhaus (1997)] and where we have set $X_{j,T} = 0$, if $j \notin \{1, \dots, T\}$. Since $I_N^X(u, \lambda)$ serves as a local estimate for the spectral density $f(u, \lambda)$, we obtain global estimates for the two integrals by

$$(2.16) \quad \hat{F}_{1,T} = \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2,$$

$$(2.17) \quad \hat{F}_{2,T} = \frac{1}{N} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \left(\frac{1}{M} \sum_{j=1}^M I_N^X(u_j, \lambda_k) \right)^2,$$

where we use the notation

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

Finally, the estimate of the measure of stationarity is given by

$$(2.18) \quad \hat{D}_T^2 = 2\pi \hat{F}_{1,T} - 4\pi \hat{F}_{2,T}.$$

In the following section we will investigate the asymptotic properties of the statistic \hat{D}_T^2 for an increasing sample size.

3 Asymptotic properties and statistical applications

In order to establish the asymptotic properties of the estimate proposed in Section 2 we require the following basic assumptions. As noted above, we have $T = NM$, and we assume that $T, M, N \rightarrow \infty$ such that

$$(3.1) \quad N \rightarrow \infty, M \rightarrow \infty, \frac{T^{1/2}}{N} \rightarrow 0, \frac{N}{T^{3/4}} \rightarrow 0.$$

Our first result specifies the asymptotic distribution of the vector $(\hat{F}_{1,T}, \hat{F}_{2,T})^T$ defined by (2.16) and (2.17).

Theorem 3.1 *If the assumptions (2.4)–(2.6) and (3.1) are satisfied, then*

$$(3.2) \quad \sqrt{T}\{(\hat{F}_{1,T}, \hat{F}_{2,T})^T - (F_1, F_2 + d_{N,T})^T\} \xrightarrow{D} \mathcal{N}(0, \Sigma),$$

where the covariance matrix Σ and the constant $d_{N,T}$ are given by

$$(3.3) \quad \Sigma = \begin{pmatrix} \frac{5}{\pi} \int_0^1 \int_{-\pi}^{\pi} f^4(u, \lambda) d\lambda du & \frac{2}{\pi} \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda \\ \frac{2}{\pi} \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda & \frac{1}{\pi} \int_{-\pi}^{\pi} \left(\left(\int_0^1 f(u, \lambda) du \right)^2 \int_0^1 f^2(u, \lambda) du \right) d\lambda \end{pmatrix}$$

and

$$(3.4) \quad d_{N,T} = \frac{N}{4\pi T} \int_0^1 \int_{-\pi}^{\pi} f^2(u, \lambda) d\lambda du,$$

respectively.

Proof. For a proof of the asymptotic normality in Theorem 3.1 we use the Cramér-Wold device and show weak convergence of the linear combination

$$(3.5) \quad A_T(c) = c^T \sqrt{T}\{(\hat{F}_{1,T}, \hat{F}_{2,T})^T - (F_1, F_2 + d_{N,T})^T\} \xrightarrow{D} \mathcal{N}(0, c^T \Sigma c)$$

for all vectors $c \in \mathbb{R}^2$, where the quantities F_1 and F_2 are defined in (2.13) and (2.14), respectively, and the matrix Σ and the constant $d_{N,T}$ are given in (3.3) and (3.4). For this purpose we show in a first step that the l th cumulant of the statistic $A_T(c)$ satisfies

$$(3.6) \quad \text{cum}_l(A_T(c)) = o(1)$$

whenever $l = 1$ or $l \geq 3$. Afterwards, we calculate the variances and covariances of $\hat{F}_{1,T}, \hat{F}_{2,T}$ and obtain

$$(3.7) \quad \lim_{T \rightarrow \infty} T \text{Var}(\hat{F}_{1,T}) = \frac{5}{\pi} \int_0^1 \int_{-\pi}^{\pi} f^4(u, \lambda) d\lambda du$$

$$(3.8) \quad \lim_{T \rightarrow \infty} T \text{Var}(\hat{F}_{2,T}) = \frac{1}{\pi} \int_{-\pi}^{\pi} \left(\left(\int_0^1 f(u, \lambda) du \right)^2 \int_0^1 f^2(u, \lambda) du \right) d\lambda$$

$$(3.9) \quad \lim_{T \rightarrow \infty} T \text{Cov}(\hat{F}_{1,T}, \hat{F}_{2,T}) = \frac{2}{\pi} \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda.$$

The assertion then follows because the cumulants of the random variable $A_T(c)$ converge to the cumulants of a normal distribution with mean 0 and variance $c^T \Sigma c$. The technical details are given in the Appendix. \square

Now a straightforward application of the Delta-method yields the asymptotic distribution of the statistic \hat{D}_T^2 defined in (2.18).

Theorem 3.2 *If the assumptions of Theorem 3.1 are satisfied, then we have*

$$\sqrt{T} \left(\hat{D}_T^2 - D^2 + 4\pi d_{N,T} \right) \xrightarrow{D} N(0, \tau^2),$$

where the constant $d_{N,T}$ is defined in (3.4) and the asymptotic variance is given by

$$(3.10) \quad \begin{aligned} \tau^2 = & 20\pi \int_0^1 \int_{-\pi}^{\pi} f^4(u, \lambda) d\lambda du - 32\pi \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda \\ & + 16\pi \int_{-\pi}^{\pi} \left(\left(\int_0^1 f(u, \lambda) du \right)^2 \int_0^1 f^2(u, \lambda) du \right) d\lambda. \end{aligned}$$

Note that there appears a non-vanishing bias

$$4\pi d_{N,T} = \frac{N}{T} \int_{-\pi}^{\pi} \int_0^1 f^2(u, \lambda) dud\lambda = \frac{2\pi N}{T} F_1$$

in Theorem 3.2, which vanishes if $N = o(\sqrt{T})$. However, this condition is excluded by the assumptions in (3.1). Nevertheless, the bias can easily be estimated by the statistic

$$B_T := \frac{2\pi N}{T^2} \sum_{j=1}^M \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2 = \frac{2\pi N}{T} \hat{F}_1^T.$$

It follows from the proof of Theorem 3.1 in the Appendix that

$$\sqrt{T} \left(B_T - 4\pi d_{N,T} \right) = \frac{2\pi N}{T} \sqrt{T} \left(\hat{F}_{1,T} - F_1 \right) \xrightarrow{P} 0,$$

and Theorem 3.2 yields

$$(3.11) \quad \sqrt{T} \left(\hat{D}_T^2 - D^2 + B_T \right) \xrightarrow{D} N(0, \tau^2).$$

For statistical applications it remains to estimate the asymptotic variance τ^2 . In general (if $D^2 > 0$), this can be accomplished by estimating the three integrals in (3.10) by rescaled versions of

$$(3.12) \quad \hat{\tau}_1^2 = \frac{1}{6T} \sum_{j=1}^M \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^4$$

$$(3.13) \quad \hat{\tau}_2^2 = \frac{2}{3NM^2} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \sum_{j_1, j_2=1}^M I_N^X(u_{j_1}, \lambda_k) I_N^X(u_{j_2}, \lambda_k)^3$$

$$(3.14) \quad \hat{\tau}_3^2 = \frac{2}{NM^3} \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \sum_{j_1, j_2, j_3=1}^M I_N^X(u_{j_1}, \lambda_k) I_N^X(u_{j_2}, \lambda_k) I_N^X(u_{j_3}, \lambda_k)^2.$$

Under the assumption of stationarity (i.e. $D^2 = 0$) the asymptotic variance in (3.10) reduces to

$$\tau_{H_0}^2 = 4\pi \int_{-\pi}^{\pi} f^4(\lambda) d\lambda,$$

and one estimates $\tau_{H_0}^2$ by $\hat{\tau}_{H_0}^2 = 4\pi^2 \hat{\tau}_1^2$. The following result shows that the statistics $\hat{\tau}_{H_0}^2$ and

$$(3.15) \quad \hat{\tau}_{H_1}^2 = 20\pi^2 \hat{\tau}_1^2 - 32\pi^2 \hat{\tau}_2^2 + 16\pi^2 \hat{\tau}_3^2$$

are consistent estimates for the asymptotic variance τ^2 in the cases $D^2 = 0$ and $D^2 > 0$, respectively. It can be shown in a similar manner as Theorem 3.1 and its proof is therefore omitted.

Theorem 3.3 *If the assumptions of Theorem 3.1 are satisfied, we have*

$$\begin{aligned} \hat{\tau}_1^2 &\xrightarrow{P} \frac{1}{\pi} \int_0^1 \int_{-\pi}^{\pi} f^4(u, \lambda) d\lambda du \\ \hat{\tau}_2^2 &\xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda \\ \hat{\tau}_3^2 &\xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} \left(\left(\int_0^1 f(u, \lambda) du \right)^2 \int_0^1 f^2(u, \lambda) du \right) d\lambda \end{aligned}$$

Remark 3.4

(a) If D^2 is used as a measure for the deviation from stationarity of a locally stationary process, we obtain from Theorem 3.2 a consistent estimate, and by Theorem 3.3 it follows that the interval

$$(3.16) \quad \left[0, \hat{D}_T^2 + B_T + \frac{\hat{\tau}_{H_1}}{\sqrt{T}} u_{1-\alpha} \right]$$

is an asymptotic $(1 - \alpha)$ confidence interval for the “parameter” D^2 , where $u_{1-\alpha}$ denotes the $(1 - \alpha)$ quantile of the standard normal distribution. The coverage probability of (3.16) in finite sample situations is investigated in Section 4.

(b) A further important application of the asymptotic results consists in the construction of an asymptotic level α test for the hypothesis of stationarity in locally stationary time series. Observing that the hypotheses (1.2) is equivalent to $D^2 = 0$ this can be accomplished by rejecting the null hypothesis whenever

$$(3.17) \quad \hat{D}_T^2 + B_T \geq \frac{\hat{\tau}_{H_0}}{\sqrt{T}} u_{1-\alpha}$$

where $\hat{\tau}_{H_0}^2$ denotes the estimate of the asymptotic variance under the null hypothesis. Moreover, the asymptotic power of this test can be approximated by a further application of Theorem 3.2, that is

$$P_{H_0}(\text{“stationarity is rejected”}) \approx \Phi\left(\sqrt{T} \frac{D^2}{\tau_{H_1}} - \frac{\tau_{H_0}}{\tau_{H_1}} u_{1-\alpha}\right),$$

where τ_{H_0} and τ_{H_1} denote the (asymptotic) standard deviation of $\sqrt{T}\hat{D}_T^2$ under the null hypothesis and alternative, respectively, and Φ is the distribution function of the standard normal distribution.

(c) Note that the results presented in this section provide an asymptotic level α test for the so called *precise hypotheses*

$$(3.18) \quad H_0 : D^2 > \varepsilon \quad \text{versus} \quad H_1 : D^2 \leq \varepsilon ,$$

[see Berger and Delampady (1987)]. The motivation for considering hypotheses of this type consists in the fact that in practice a (locally stationary) time series will usually never be precisely stationary, and a more realistic question in this context would be, if the process shows approximately stationary behavior [see also the discussion in Remark 3.6]. Therefore the parameter $\varepsilon > 0$ in (3.18) denotes a prespecified constant for which the statistician agrees to analyse the data under the additional assumption of stationarity. An asymptotic level α test for the hypothesis (3.18) is obtained by rejecting the null hypothesis, whenever

$$(3.19) \quad \hat{D}_T^2 - \varepsilon + B_T < \frac{\hat{\tau}_{H_1}}{\sqrt{T}} u_\alpha .$$

Note that this procedure allows for accepting the null hypothesis of “approximate stationarity” at controlled type I error.

Remark 3.5 It should be noted that the results in Theorem 3.2 can be extended to the case where the innovations are not necessarily normal distributed. This assumption simplifies the argument in the proof substantially but can be weakened to the case of independent identically distributed random variables with existing moments of all order. In this general case Theorem 3.2 remains valid with a different asymptotic variance, i.e.

$$\begin{aligned} \tau_g^2 = & 20\pi \int_0^1 \int_{-\pi}^\pi f^4(u, \lambda) d\lambda du - 32\pi \int_{-\pi}^\pi \left(\int_0^1 f(u, \lambda) du \int_0^1 f^3(u, \lambda) du \right) d\lambda \\ & + 16\pi \int_{-\pi}^\pi \left(\left(\int_0^1 f(u, \lambda) du \right)^2 \int_0^1 f^2(u, \lambda) du \right) d\lambda \\ & + \frac{\kappa_4}{\kappa_2^2} \left\{ 4 \int_0^1 \left(\int_{-\pi}^\pi f^2(u, \lambda) d\lambda \right)^2 du - 8 \int_0^1 \left(\int_{-\pi}^\pi f(u, \lambda)^2 d\lambda \int_{-\pi}^\pi f(u, \lambda) \left(\int_0^1 f(v, \lambda) dv \right) d\lambda \right) du \right. \\ & \left. + 4 \int_0^1 \left(\int_{-\pi}^\pi f(u, \lambda) \left(\int_0^1 f(v, \lambda) dv \right) d\lambda \right)^2 du \right\} \end{aligned}$$

where κ_2 and κ_4 denote the variance and the fourth cumulant of the innovations. Therefore, even though τ_g^2 is in general different from τ^2 , both quantities coincide at least in the stationary case.

Remark 3.6 Following Dahlhaus (1997) it is too restrictive to use the more natural definition

$$(3.20) \quad X_{t,T} = \sum_{l=-\infty}^{\infty} \psi_l(t/T) Z_{t-l}, \quad t = 1, \dots, T,$$

for a locally stationary process, as in this case even time-varying AR(1)-processes are ruled out. This explains the need for the more general class of processes introduced in (2.1). As a drawback, the spectral density function has to be defined via the approximating sequence ψ_l , and this means in particular that even $f(u, \lambda) = f(\lambda)$ does not imply stationarity of $X_{t,T}$, as one can only conclude that the time varying coefficients $\psi_{l,t,T}$ can be approximated by constants ψ_l . Thus the minimal distance D^2 formally plays the role of a best approximation of the time-varying spectral density by a time-homogeneous function, but to avoid confusion we still refer to this case as the stationary one. This concept is standard in the context of investigating stationarity in locally stationary processes [see for example Paparoditis (2009, 2010) or Dahlhaus (2009)].

4 Finite sample properties

In this section we study the finite sample properties of the asymptotic confidence intervals for the parameter D^2 and of the test for stationarity. All results are based on 1000 simulation runs.

4.1 Confidence intervals

The coverage probability of the confidence intervals defined in (3.16) is investigated for the tvMA(2) model

$$(4.1) \quad X_{t,T} = 2Z_t - \left\{ 1 + b \cos\left(2\pi \frac{t}{T}\right) \right\} Z_{t-1},$$

where different choices for the parameter b are considered and the Z_t are independent, standard Gaussian distributed random variables. Note that the condition (3.1) implies

$$(4.2) \quad N = O(T^\beta) \text{ with } \beta \in (1/2, 3/4),$$

and we recommend to choose the parameter M sufficiently large in order to account for the local structure of the time series in a satisfying way. The results are displayed in Table 1 for various values of T and M (which determines N). We observe reasonable coverage probabilities in the cases $b = 0$ and $b = 0.5$. In the case of stationarity ($b = 0$), the actual coverage probability is in fact larger than the pre-specified level, while the opposite behavior is observed in the case $b = 1$. Based on our numerical experiments, we conclude that the choice $M = 16$ is sufficient for most of the examples while in some cases $M = 32$ leads to better results [see the part corresponding to $b = 1$ in Table 1]. Table 1 shows that the coverage probability is satisfying even for smaller values of M if $b = 0.5$ or $b = 0$. An intuitive explanation for these observations is that a smaller value of b yields a lower time-dependency of the spectral density, and as a consequence, smaller values of the factor M are required for efficient analysis.

4.2 Testing for stationarity

We next study the size and the power of the test in (3.17) by calculating the rejection frequencies for different values of b in the model (4.1). The corresponding results are displayed in Table 2. Note that

| | | | $b = 0$ | | $b = 0.5$ | | $b = 1$ | |
|------|-----|-----|---------|-------|-----------|-------|---------|-------|
| T | N | M | 95% | 90% | 95% | 90% | 95% | 90% |
| 256 | 32 | 8 | 0.993 | 0.968 | 0.968 | 0.901 | 0.828 | 0.760 |
| 512 | 64 | 8 | 0.983 | 0.941 | 0.958 | 0.908 | 0.838 | 0.784 |
| 1024 | 64 | 16 | 0.966 | 0.925 | 0.944 | 0.906 | 0.884 | 0.833 |
| 1024 | 128 | 8 | 0.976 | 0.935 | 0.944 | 0.886 | 0.829 | 0.764 |
| 2048 | 64 | 32 | 0.978 | 0.949 | 0.960 | 0.931 | 0.911 | 0.882 |
| 2048 | 128 | 16 | 0.945 | 0.910 | 0.919 | 0.872 | 0.873 | 0.837 |

Table 1: Coverage Probability of the asymptotic confidence interval (3.16) in 1000 replications of the $tvMA(2)$ model (4.1) for different values of b .

| | | | $H_0 : b = 0$ | | $H_1 : b = 0.5$ | | $H_1 : b = 1$ | |
|------|-----|-----|---------------|-------|-----------------|-------|---------------|-------|
| T | N | M | 5% | 10% | 5% | 10% | 5% | 10% |
| 256 | 32 | 8 | 0.062 | 0.138 | 0.090 | 0.200 | 0.266 | 0.463 |
| 512 | 64 | 8 | 0.053 | 0.125 | 0.117 | 0.234 | 0.430 | 0.582 |
| 1024 | 64 | 16 | 0.079 | 0.166 | 0.224 | 0.381 | 0.758 | 0.866 |
| 1024 | 128 | 8 | 0.038 | 0.122 | 0.140 | 0.263 | 0.690 | 0.825 |

Table 2: Rejection probabilities of the test (3.17) in 1000 replications of the $tvMA(2)$ model (4.1) for different values of b .

under the null hypothesis (i.e. $b = 0$) the spectral density does not depend on u so that even small values of M lead to a precise approximation of the nominal level of the test. Also, for larger values of N the approximation of the nominal level is more accurate, whereas on the other hand a larger M leads to a more satisfying power behaviour. To obtain both a reasonable approximation of the level and a good behavior of the test, we therefore recommend to choose β [see (4.2)] in the middle of the interval $(1/2, 3/4)$, i.e. close to $5/8$, and for this choice we observe that the test yields reasonable rejection probabilities under the alternative ($b = 0.5, 1.0$).

4.3 Validating stationarity

Finally, we investigate the test for the precise hypothesis (3.18) proposed in (3.19) where the bound for accepting stationarity is chosen as $\varepsilon = 0.9$. Note that for the values $b = 1$, $b = 0.5$ and $b = 0$ we obtain $D_{b=1}^2 \approx 0.972$, $D_{b=0.5}^2 \approx 0.239$ and $D_{b=0}^2 = 0$, respectively. Therefore the cases $b = 0$ and $b = 0.5$ correspond to the alternative $H_1 : D^2 < 0.9$, while the choice $b = 1$ gives a scenario for the

| | | | $H_1 : b = 0$ | | $H_1 : b = 0.5$ | | $H_0 : b = 1$ | |
|------|-----|-----|---------------|-------|-----------------|-------|---------------|-------|
| T | N | M | 5% | 10% | 5% | 10% | 5% | 10% |
| 256 | 32 | 8 | 0.421 | 0.526 | 0.348 | 0.442 | 0.142 | 0.209 |
| 512 | 64 | 8 | 0.677 | 0.763 | 0.464 | 0.547 | 0.131 | 0.185 |
| 1024 | 64 | 16 | 0.614 | 0.69 | 0.515 | 0.608 | 0.096 | 0.127 |
| 1024 | 128 | 8 | 0.841 | 0.88 | 0.625 | 0.697 | 0.147 | 0.193 |
| 2048 | 64 | 32 | 0.825 | 0.873 | 0.559 | 0.632 | 0.064 | 0.082 |
| 2048 | 128 | 16 | 0.779 | 0.847 | 0.709 | 0.780 | 0.091 | 0.132 |

Table 3: *Rejection frequencies of the test (3.19) for the precise hypothesis (3.18) in 1000 replications of the tvMA(2) model (4.1) for different values of b .*

null hypothesis $H_0 : D^2 \geq 0.9$. The results are depicted in Table 3. As in Section 4.1, we recommend to choose $M \geq 16$ to obtain a satisfying size of the test. In these cases the level of the test is usually overestimated. This observation can be explained by the fact that for the choice $b = 1$ there is a strong deviation from stationarity, and consequently, a reasonable sample size is required in order to obtain a precise approximation of the nominal level.

4.4 Data examples

In this subsection we illustrate the application of the developed methodology by re-analyzing several data examples from the recent literature. We begin with an example from neuroscience which has been considered in von Sachs and Neumann (2000) and Paparoditis (2009). These authors analyzed a data set of tremor data recorded in the Cognitive Neuroscience Laboratory of the University of Québec at Montreal. There are 3071 observations and the purpose of the study is a comparison of different regions of tremor activity coming from a subject with Parkinson’s disease. In the left part of Figure 2 we show a plot of the estimate

$$(4.3) \quad \hat{f}(u, \lambda) = \frac{2\pi}{N} \sum_{j=1}^N \frac{1}{b} K\left(\frac{\lambda - \frac{2\pi j}{N}}{b}\right) I_N^X\left(u, \frac{2\pi j}{N}\right)$$

for the two dimensional density $f(u, \lambda)$, where $N = 256$ and $b = 0.18$ [see Paparoditis (2009) for a similar approach]. The plot indicates some non stationarity in the data and it might be of interest to investigate this visual conclusion by the statistical methodology developed in this paper. For the calculation of the test statistic we used $N = 192$ and $M = 16$ in order to address for non stationary behavior of the time series and to keep the bias reasonably small. For the measure D^2 of stationarity we obtain $\hat{D}^2 = \hat{D}_T^2 + B_T \approx 3.56 \times 10^{-7}$ with a standard deviation of $\hat{\tau}_{H_1} \approx 1.06 \times 10^{-5}$. This yields for the standardized distance $\sqrt{T} \frac{D^2}{\tau}$ the estimate $\sqrt{3071} \frac{\hat{D}^2}{\hat{\tau}_{H_1}} \approx 1.884$ and the test for stationarity rejects

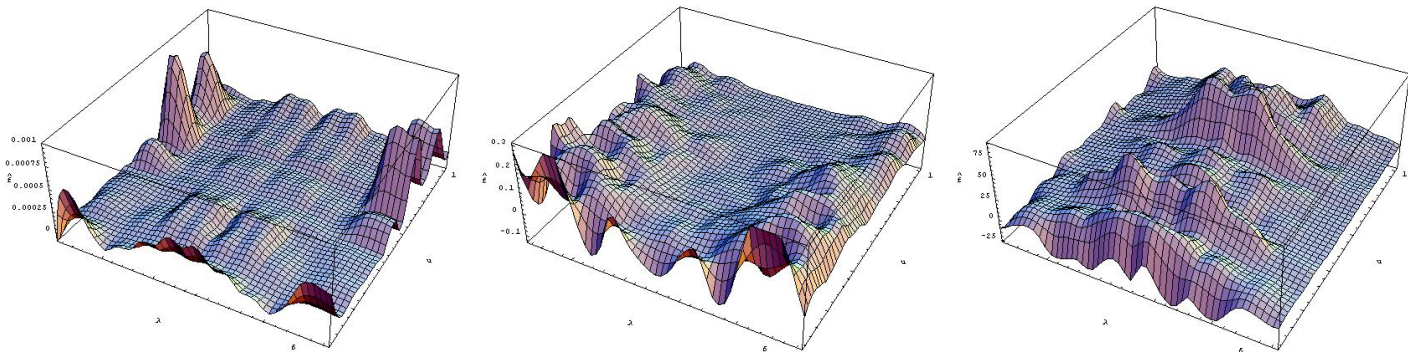


Figure 2: Estimate of the spectral density for three data sets. Left panel: neuroscience data. Middle panel: egg data. Right panel: infant data.

the null hypothesis with a p -value of 0.033. Note that these findings confirm the investigations in Paparoditis (2009).

In our second example we investigate 1201 observations of weekly egg prices at a German agriculture market between April 1967 and May 1990. Following Paparoditis (2010) the first-order differences $\Delta_t = X_t - X_{t-1}$ of the observed time series are analyzed. For the calculation of the estimate \hat{D}^2 and the test statistic we chose $N = 80$, $M = 15$ and obtain $\hat{D}^2 = \hat{D}_T^2 + B_T \approx 0.0013$, $\hat{\tau}_{H_1} \approx 0.0967$, which yields for the standardized distance $\sqrt{T} \frac{D^2}{\tau}$ the estimate $\sqrt{1200} \frac{\hat{D}^2}{\hat{\tau}_{H_1}} \approx 0.454$. A plot of the density estimate (4.3) is shown in middle panel of Figure 2, where we used $\hat{N} = 134$ and $b = 0.112$. Although this plot shows some non stationary behavior for small and large values of u we obtain a p -value of 0.321 and the null hypothesis of stationarity cannot be rejected. These observations are different to the result obtained by Paparoditis (2010). An explanation could be that we smooth the differences between the local spectral density and the best stationary approximation over time while Paparoditis (2010) takes the maximum as the test statistic.

In our final example we re-analyze a time series, which shows the heart rate electrocardiogram (ECG) of an 66-day-old infant, sampled at 1/16 Hz and recorded from 21:17:59 to 6:27:18 leading to 2048 observations. This data set was also considered by von Sachs and Neumann (2000) and Paparoditis (2010). We investigated the first-order differences and the plot of the estimate of the local spectral density $\hat{f}(u, \lambda)$ clearly indicates a non stationary behavior (here we use $N = 194$ and $b = 0.095$). We have applied the methodology developed in this paper with $N = 128$, $M = 16$ and obtain $\hat{D}^2 = \hat{D}_T^2 + B_T \approx 478.994$ and $\hat{\tau}_{H_1} \approx 10485.47$. This yields for the standardized distance $\sqrt{T} \frac{D^2}{\tau}$ the estimate $\sqrt{2047} \frac{\hat{D}^2}{\hat{\tau}_{H_1}} \approx 2.081$. The test for stationarity rejects the null hypothesis with a p -value of 0.022. These results confirm the findings of Paparoditis (2010).

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research

Foundation (DFG).

5 Appendix: technical details

From (2.3) it follows that the error due to the approximation of $X_{t,T}$ by the corresponding series from (3.20) is of small order. Therefore we use the latter representation without further mentioning.

5.1 Proof of the estimate (3.6) in the case $l = 1$

In order to prove (3.6) in the case $l = 1$ we can treat the statistics $\hat{F}_{1,T}$ and $\hat{F}_{2,T}$ separately because of the linearity of the expectation. For the sake of brevity, we restrict ourselves to the (more complicated) statement for the statistic $\sqrt{T}(\hat{F}_{1,T} - F_1)$ and prove

$$(5.1) \quad \mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^M \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2 - \sqrt{T} \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} f^2(u, \lambda) d\lambda du \right) = o(1).$$

In order to show this estimate we introduce some additional notation. We set $t_{j,p} = t_j - N/2 + 1 + p$ and $u_{j,p} = t_{j,p}/T$ and obtain

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2 \right) \\ &= \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p - q + r - s)) \\ & \quad \times \psi_l(u_{j,p}) \psi_m(u_{j,q}) \psi_n(u_{j,r}) \psi_o(u_{j,s}) \mathbb{E}(Z_{t_{j,p-l}} Z_{t_{j,q-m}} Z_{t_{j,r-n}} Z_{t_{j,s-o}}) + O\left(\frac{1}{T}\right). \end{aligned}$$

Since Z is normally distributed by assumption, we have

$$\mathbb{E}(Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4}) = \mathbb{E}(Z_{i_1} Z_{i_2}) \mathbb{E}(Z_{i_3} Z_{i_4}) + \mathbb{E}(Z_{i_1} Z_{i_3}) \mathbb{E}(Z_{i_2} Z_{i_4}) + \mathbb{E}(Z_{i_1} Z_{i_4}) \mathbb{E}(Z_{i_2} Z_{i_3})$$

for arbitrary indices i_1, i_2, i_3, i_4 , and thus we obtain by means of a Taylor expansion

$$(5.2) \quad \mathbb{E} \left(\frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2 \right) = E_{N,T}^1 + E_{N,T}^2 + E_{N,T}^3 + A_{N,T} + O\left(\frac{1}{T}\right) + O\left(\frac{N^2}{T^2}\right),$$

where we set

$$E_{N,T}^1 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p-q+r-s)) \psi_l(u_j) \psi_m(u_j)$$

$$\times \psi_n(u_j) \psi_o(u_j) \mathbb{E}(Z_{t_j,p-l} Z_{t_j,q-m}) \mathbb{E}(Z_{t_j,r-n} Z_{t_j,s-o})$$

$$E_{N,T}^2 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p-q+r-s)) \psi_l(u_j) \psi_m(u_j)$$

$$\times \psi_n(u_j) \psi_o(u_j) \mathbb{E}(Z_{t_j,p-l} Z_{t_j,s-o}) \mathbb{E}(Z_{t_j,q-m} Z_{t_j,r-n})$$

$$E_{N,T}^3 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p-q+r-s)) \psi_l(u_j) \psi_m(u_j)$$

$$\times \psi_n(u_j) \psi_o(u_j) \mathbb{E}(Z_{t_j,p-l} Z_{t_j,r-n}) \mathbb{E}(Z_{t_j,q-m} Z_{t_j,s-o})$$

$$A_{N,T} = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p-q+r-s))$$

$$\times \left\{ \left(\psi_l(u_{j,p}) - \psi_l(u_j) \right) \psi_m(u_j) \psi_n(u_j) \psi_o(u_j) + \psi_l(u_j) \left(\psi_m(u_{j,q}) - \psi_m(u_j) \right) \psi_n(u_j) \psi_o(u_j) \right. \\ \left. + \psi_l(u_j) \psi_m(u_j) \left(\psi_n(u_{j,r}) - \psi_n(u_j) \right) \psi_o(u_j) + \psi_l(u_j) \psi_m(u_j) \psi_n(u_j) \left(\psi_o(u_{j,s}) - \psi_o(u_j) \right) \right\} \\ \times \mathbb{E}(Z_{t_j,p-l} Z_{t_j,q-m} Z_{t_j,r-n} Z_{t_j,s-o}).$$

The terms $E_{N,T}^1$, $E_{N,T}^2$, $E_{N,T}^3$ and $A_{N,T}$ are now treated separately showing

$$(5.3) \quad E_{N,T}^1 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} f^2(u_j, \lambda_k) + O\left(\frac{1}{N}\right)$$

$$(5.4) \quad E_{N,T}^2 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} f^2(u_j, \lambda_k) + O\left(\frac{1}{N}\right)$$

$$(5.5) \quad E_{N,T}^3 = O(1/N)$$

$$(5.6) \quad A_{N,T} = O\left(\frac{N^2}{T^2}\right) + O\left(\frac{1}{T}\right).$$

The estimates (5.3) and (5.4) follow by similar arguments and we restrict ourselves to a proof of the

first one. A standard calculation observing that $E(Z_{i_1} Z_{i_2}) = 0$ for $i_1 \neq i_2$ shows that

$$\begin{aligned} E_{N,T}^1 &= \frac{\sigma^4}{T(2\pi N)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(l-m+n-o)) \\ &\quad \times \psi_l(u_j) \psi_m(u_j) \psi_n(u_j) \psi_o(u_j) \max(0, N-|l-m|) \max(0, N-|n-o|) \\ &= \frac{\sigma^4}{T(2\pi)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{\substack{l,m,n,o=-\infty \\ |l-m| \leq N-1 \\ |n-o| \leq N-1}}^{\infty} \exp(-i\lambda_k(l-m+n-o)) \psi_l(u_j) \psi_m(u_j) \psi_n(u_j) \psi_o(u_j) + O\left(\frac{1}{N}\right), \end{aligned}$$

where we have used (2.4) in the last equality. With (2.4) it also follows that

$$(5.7) \quad \sum_{|l| \geq N} \sup_u |\psi_l(u)| = O(1/N^2),$$

and then it is easy to see that we can drop the restrictions on $|l-m|$ and $|n-o|$ in the summation as well. Therefore $E_{N,T}^1$ reduces to

$$E_{N,T}^1 = \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} f^2(u_j, \lambda_k) + O\left(\frac{1}{N}\right),$$

which proves (5.3). Similarly, observing (5.7) and the identity

$$(5.8) \quad \sum_{r=0}^{N-1} \exp(-i2\lambda_k r) = \begin{cases} N, & k = 0 \text{ or } k = \frac{N}{2} \\ 0, & \text{else} \end{cases}$$

we obtain by a straightforward but tedious calculation

$$\begin{aligned} E_{N,T}^3 &= \frac{\sigma^4}{T(2\pi N)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{l,m,n,o=-\infty}^{\infty} \psi_l(u_j) \psi_m(u_j) \psi_n(u_j) \psi_o(u_j) \\ &\quad \times \sum_{\substack{r,s=0 \\ 0 \leq r+l-n \leq N-1 \\ 0 \leq s+m-o \leq N-1}}^{N-1} \exp(-i\lambda_k(2r-2s+l-n-m+o)) \\ &= \frac{\sigma^4}{T(2\pi N)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{l,m,n,o=-\infty}^{\infty} \psi_l(u_j) \psi_m(u_j) \psi_n(u_j) \psi_o(u_j) \\ &\quad \times \sum_{r,s=0}^{N-1} \exp(-i\lambda_k(2r-2s+l-n-m+o)) + O(1/N^2) \\ &= \frac{1}{TN^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} f^2(u_j, \lambda_k) \left| \sum_{r=0}^{N-1} \exp(-i2\lambda_k r) \right|^2 + O(1/N^2) \\ &= O(1/N), \end{aligned}$$

which proves (5.5). Finally, for the estimate of the term $A_{N,T}$ we use a Taylor expansion and the condition $\sup_{l \in \mathbb{R}} \sup_{u \in [0,1]} |\psi_l''(u)| < \infty$ to obtain [note that $u_{j,p} - u_j = (p + 1 - N/2)/T$]

$$\begin{aligned}
A_{N,T} &= \frac{2}{T(2\pi N)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p - q + r - s)) \\
&\quad \times \left[\left(\psi_l(u_{j,p}) - \psi_l(u_j) \right) \psi_m(u_j) + \psi_l(u_j) \left(\psi_m(u_{j,q}) - \psi_m(u_j) \right) \right] \psi_n(u_j) \psi_o(u_j) \\
&\quad \times \mathbb{E}(Z_{t_j,p-l} Z_{t_j,q-m} Z_{t_j,r-n} Z_{t_j,s-o}) \\
&= \frac{2}{T(2\pi N)^2} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{p,q,r,s=0}^{N-1} \sum_{l,m,n,o=-\infty}^{\infty} \exp(-i\lambda_k(p - q + r - s)) \\
&\quad \times \left[\psi_l'(u_j) \psi_m(u_j)^{\frac{p+1-N/2}{T}} + \psi_l(u_j) \psi_m'(u_j)^{\frac{q+1-N/2}{T}} \right] \psi_n(u_j) \psi_o(u_j) \\
&\quad \times \mathbb{E}(Z_{t_j,p-l} Z_{t_j,q-m} Z_{t_j,r-n} Z_{t_j,s-o}) + O\left(\frac{N^2}{T^2}\right).
\end{aligned}$$

This expression is now treated by similar arguments as given for the terms $E_{n,T}^1$ and $E_{n,T}^3$, which yields the estimate (5.6).

The assertion (3.6) in the case $l = 1$ now follows from (5.3)–(5.6) and by assumption on the growth of N and T , observing that the sums in (5.3) and (5.4) can be approximated by the integrals

$$\frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} f^2(u, \lambda) d\lambda du$$

with an error of order $O\left(\frac{1}{N^2} + \frac{N^2}{T^2}\right)$ each, which is due to the choice of the midpoints u_j and λ_k and to the periodicity of f in its second component.

5.2 Calculation of the cumulants in the case $l \geq 2$

In this final part of the proof we show the convergence of the l th cumulant ($l \geq 2$) of the statistic $A_T(c)$ towards that of the normal distribution specified in (3.5). We start with the claim

$$(5.9) \quad \text{cum}_l(A_T(c)) = O(T^{1-l/2}) \quad \text{for } l \geq 2,$$

which shows, inter alia, that the cumulants of degree higher than two converge to zero as requested. Throughout the proof we restrict ourselves to the case $c = (1, 0)$, as the general one follows from exactly the same lines with an additional amount of notation.

As before it suffices to show the assertion for the l th cumulant of

$$\begin{aligned}
& \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_j, \lambda_k)^2 \\
&= \frac{1}{T} \sum_{j=1}^M \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(2\pi N)^2} \sum_{p,q,r,s=0}^{N-1} \sum_{v,w,x,y=-\infty}^{\infty} \exp(-i\lambda_k(p-q+r-s)) \\
&\quad \times \psi_v(u_{j,p}) \psi_w(u_{j,q}) \psi_x(u_{j,r}) \psi_y(u_{j,s}) \mathbb{E}(Z_{t_{j,p-v}} Z_{t_{j,q-w}} Z_{t_{j,r-x}} Z_{t_{j,s-y}}) + O\left(\frac{1}{T}\right)
\end{aligned}$$

after rescaling. Using the multilinearity of the cumulant and the product theorem for cumulants [see Brillinger (1981) and its terminology] that quantity becomes

$$(5.10) \quad T^{l/2} \text{cum} \left(\frac{1}{T} \sum_{j_1=1}^M \sum_{k_1=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_{j_1}, \lambda_{k_1})^2, \dots, \frac{1}{T} \sum_{j_l=1}^M \sum_{k_l=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} I_N^X(u_{j_l}, \lambda_{k_l})^2 \right) = \frac{1}{(2\pi)^{2l}} \sum_{\nu} V(\nu).$$

In order to define $V(\nu)$ we have to use some further notation. First we introduce

$$Y_{i1} = Z_{t_{j_i, p_i - v_i}}, \quad Y_{i2} = Z_{t_{j_i, q_i - w_i}}, \quad Y_{i3} = Z_{t_{j_i, r_i - x_i}}, \quad Y_{i4} = Z_{t_{j_i, s_i - y_i}}$$

for $i \in \{1, \dots, l\}$. Then we set

$$\begin{aligned}
(5.11) \quad V(\nu) &= \frac{1}{T^{l/2}} \frac{1}{N^{2l}} \sum_{v_1, \dots, v_l = -\infty}^{\infty} \sum_{j_1, \dots, j_l = 1}^M \sum_{p_1, \dots, s_l = 0}^{N-1} \sum_{k_1, \dots, k_l = -\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \psi_{v_1}(u_{j_1, p_1}) \dots \psi_{v_l}(u_{j_l, s_l}) \\
&\quad \times \exp(-i\lambda_{k_1}(p_1 - q_1 + r_1 - s_1)) \dots \exp(-i\lambda_{k_l}(p_l - q_l + r_l - s_l)) \\
&\quad \times \text{cum}(Y_{ik}; ik \in \nu_1) \dots \text{cum}(Y_{ik}; ik \in \nu_{2l}),
\end{aligned}$$

where the summation in (5.10) is performed with respect to all indecomposable partitions $\nu = \nu_1 \cup \dots \cup \nu_{2l}$ with subsets containing two elements (due to the normality of Z) of the table

$$(5.12) \quad \begin{array}{cccc} (1, 1) & (1, 2) & (1, 3) & (1, 4) \\ \vdots & \vdots & \vdots & \vdots \\ (l, 1) & (l, 2) & (l, 3) & (l, 4) \end{array}$$

As the number C_l of indecomposable partitions does not depend on T , it suffices to prove that each $V(\nu)$ has the desired properties. Thus we keep ν fixed. Also as ν is indecomposable, we know that each row of the table communicates with any other one, and thus we can assume without loss of generality that the i th row hooks with the $(i+1)$ th one (otherwise we switch the rows accordingly).

Let us also fix v_1, \dots, v_l and j_1 . That the first row hooks with the second one means that a product of the form $\text{cum}(Y_{11}, Y_{23})$ appears within (5.11). In order for it to be non-zero the corresponding indices of Z have to be equal, that is there has to exist a relation of the form

$$(5.13) \quad t_{j_1} - N/2 + 1 + p_1 - v_1 = t_{j_2} - N/2 + 1 + r_2 - x_2 \Leftrightarrow r_2 = p_1 - v_1 + x_2 + t_{j_1} - t_{j_2}.$$

Thus r_2 has to satisfy both

$$x_2 - v_1 + t_{j_1} - t_{j_2} \leq r_2 \leq x_2 - v_1 + t_{j_1} - t_{j_2} + N - 1 \quad \text{and} \quad 0 \leq r_2 \leq N - 1,$$

and since v_1 and x_2 are kept fixed and as $t_{j_1} - t_{j_2} = mN$ for $m \in \mathbb{Z}$, we conclude that there are at most two options for t_{j_2} (and thus for j_2) that lead to a non-zero cumulant. By induction it follows that given j_1 there is only a finite number D_l of valid choices for the indices j_2, \dots, j_l , and in the following we keep one of these fixed as well.

We have already seen in (5.13) that there are $2l$ conditions of the form

$$(5.14) \quad p_1 - r_2 = v_1 - x_2 + t_{j_2} - t_{j_1}$$

that have to be satisfied in order for the cumulants to be non-zero. Since ν is a partition, each variable p_1, \dots, s_l appears exactly once within these $2l$ expressions. Also, we know essentially from (5.8) that further l equations

$$(5.15) \quad p_i - q_i + r_i - s_i = m_i N \quad \text{with} \quad m_i \in \mathbb{Z}$$

have to be valid as well, and it is obvious that only $m_i \in \{-1, 0, 1\}$ is possible. Fix one of the E_l possible sequences m_1, \dots, m_l . In the following we will prove that the solution space of the previous system of $3l$ equations in $4l$ variables is at most of dimension $l + 1$. For this assertion it suffices to show that the solution space of the corresponding homogeneous system has the same properties.

To this end we identify \mathbb{R}^{4l} with the variables $p_1, q_1, r_1, s_1, \dots, p_l, q_l, r_l, s_l$ in that particular order. Then we set

$$(5.16) \quad v_i = (0 \ \cdots \ 0 \ 1 \ -1 \ 1 \ -1 \ 0 \ \cdots \ 0)^T \in \mathbb{R}^{4l} \quad \text{for} \quad i \in \{1, \dots, l\}$$

and

$$(5.17) \quad w_i = (0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0 \ -1 \ 0 \ \cdots \ 0)^T \in \mathbb{R}^{4l} \quad \text{for} \quad i \in \{1, \dots, 2l\},$$

where the vectors v_i and w_i relate to the homogeneous versions of the equations in (5.15) and (5.14) in an obvious way: v_i refers to the conditions involving p_i, \dots, s_i , whereas w_i represents the $2l$ equations from (5.14) in arbitrary order. The claim on the dimension of the solution space can be deduced from the following lemma.

Lemma 5.1 *The vectors $v_2, v_3, \dots, v_l, w_1, \dots, w_{2l}$ are linearly independent.*

Proof. Suppose there are constants $\alpha_2, \alpha_3, \dots, \alpha_l, \beta_1, \dots, \beta_{2l}$ such that

$$(5.18) \quad \alpha_2 v_2 + \alpha_3 v_3 + \dots + \alpha_l v_l + \beta_1 w_1 + \dots + \beta_{2l} w_{2l} = 0.$$

Focus on those w_{i_1} with a non-zero entry among the first four rows. Since v_1 is not included in the sum, the corresponding coefficients β_{i_1} have to be zero, as otherwise (5.18) would not be satisfied. Now

the partition is chosen in such a way that first row of the table hooks with the second one, thus there is a vector $w_{i_{12}}$ with one non-zero entry within rows 1 to 4 and the second non-zero entry within rows 5 to 8. As $\beta_{i_{12}}$ is zero, the same argument as before forces α_2 to be zero. The claim now follows by induction, as we have $\alpha_j = 0$, thus each $\beta_{i_j} = 0$, and the j th row hooks with the $(j + 1)$ st. \square

With the aid of these results the proof of assertion (5.9) is now easy. From the previous discussion we know that the sum in (5.11) has the following upper bound

$$|V(\nu)| \leq D_l E_l \sigma^{4l} \frac{1}{T^{l/2}} \frac{1}{N^{2l}} \sum_{v_1, \dots, v_l = -\infty}^{\infty} M N^{l+1} N^l \sup_u |\psi_{v_1}(u)| \dots \sup_u |\psi_{v_l}(u)| = O(T^{1-l/2}).$$

To complete the proof of Theorem 3.1 it remains to show the asymptotic representations (3.7)–(3.9) for the variances and covariances of the statistics \hat{F}_1^T and \hat{F}_2^T . For the sake of brevity, we restrict ourselves to the case (3.7). All other cases are treated similarly. In order to prove that claim we use (5.9) for $l = 2$, a Taylor expansion and the same notation and arguments as before to obtain

$$4T \text{Var}(\hat{F}_{1,T}) = \frac{1}{(2\pi)^4} \sum_{\nu} V(\nu) + O\left(\frac{N}{T}\right) + O\left(\frac{1}{N}\right),$$

where

$$(5.19) \quad V(\nu) = \frac{1}{T} \frac{1}{N^4} \sum_{j_1, j_2=1}^M \sum_{k_1, k_2=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \sum_{v_1, w_1, x_1, y_1=-\infty}^{\infty} \sum_{v_2, w_2, x_2, y_2=-\infty}^{\infty} \sum_{p_1, q_1, r_1, s_1=0}^{N-1} \sum_{p_2, q_2, r_2, s_2=0}^{N-1} \\ \psi_{v_1}(u_{j_1}) \psi_{w_1}(u_{j_1}) \psi_{x_1}(u_{j_1}) \psi_{y_1}(u_{j_1}) \psi_{v_2}(u_{j_2}) \psi_{w_2}(u_{j_2}) \psi_{x_2}(u_{j_2}) \psi_{y_2}(u_{j_2}) \\ \times \exp(-i\lambda_{k_1}(p_1 - q_1 + r_1 - s_1)) \exp(-i\lambda_{k_2}(p_2 - q_2 + r_2 - s_2)) \\ \times \text{cum}(Y_{ik}; ik \in \nu_1) \dots \text{cum}(Y_{ik}; ik \in \nu_p).$$

The main idea now is to single out those partitions ν for which $V(\nu)$ is of order one. Following the proof of (5.9) we know that this is not the case, if all v_i and w_i are linearly independent, as in this situation $V(\nu)$ has order $1/N$. To obtain all indecomposable partitions $\nu = \nu_1 \cup \dots \cup \nu_4$ of

$$\begin{aligned} &(1, 1) \quad (1, 2) \quad (1, 3) \quad (1, 4) \\ &(2, 1) \quad (2, 2) \quad (2, 3) \quad (2, 4) \end{aligned}$$

that lead to linearly dependent vectors, one distinguishes two cases: Either there exists exactly one set of the partition which consists of two elements of the first row (and thus there exists another set within ν that contains two elements of the second one), or in each set of ν there is one element from the first row and one from the second row.

Fix a partition ν' of the first type. One finds easily that its corresponding vectors v_i and w_i are linearly dependent, if and only if those $(1, i)$ and $(1, j)$ (and $(2, i)$ and $(2, j)$) hook, for which i is even and j is odd (and vice versa). In total there are 32 such partitions. A partition ν'' of the second type falls

into the class of interest, if either both odd components of the first row hook either with both odd components or with both even components of the second one, which gives 8 possible choices. The remainder of the proof consists of a tedious computation of $V(\nu')$ and $V(\nu'')$. It turns out that one obtains

$$V(\nu') = V(\nu'') = (2\pi)^3 \int_0^1 \int_{-\pi}^{\pi} f(u, \lambda)^4 d\lambda du + O\left(\frac{1}{N}\right) + O\left(\frac{N}{T}\right),$$

from which we conclude

$$4T \operatorname{Var}(\hat{F}_{1,T}) = \frac{40}{(2\pi)^4} V(\nu') + O\left(\frac{N}{T}\right) + O\left(\frac{1}{N}\right) = \frac{20}{\pi} \int_0^1 \int_{-\pi}^{\pi} f(u, \lambda)^4 d\lambda du + O\left(\frac{1}{N}\right) + O\left(\frac{N}{T}\right),$$

which completes the proof of Theorem 3.1. □

References

- Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93:1488–1501.
- Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2:317–335.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw Hill, New York.
- 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. (2000). A likelihood approximation for locally stationary processes. *Annals of Statistics*, 28(6):1762–1794.
- Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics*, 151:101–112.
- 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.
- Lee, S., Ha, J., Na, O., and S., N. (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics*, 30(4):781–796.

- Nason, G. P., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society, Ser. B*, 62:271–292.
- 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.
- 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.
- Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Ser. B*, 62:204–237.
- Sakiyama, K. and Taniguchi, M. (2003). Testing composite hypotheses for locally stationary processes. *Journal of Time Series Analysis*, 24(4):483–504.
- Sakiyama, K. and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis*, 90:282–300.
- Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semiparametric hypotheses for locally stationary processes. *Scandinavian Journal of Statistics*, 36:800–821.
- von Sachs, R. and Neumann, M. H. (2000). A wavelet-based test for stationarity. *Journal of Time Series Analysis*, 21:597–613.

