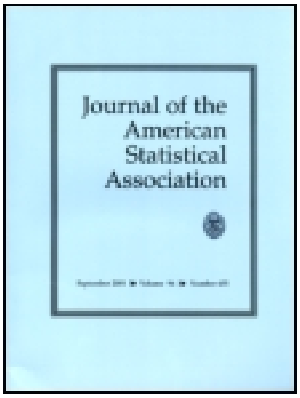

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

A Measure of Stationarity in Locally Stationary Processes With Applications to Testing

Holger Dette[a], Philip Preuß[a] & Mathias Vetter[a]

[a] Holger Dette is Professor, Philip Preuss is Ph.D. Student, and Mathias Vetter is PostDo, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany. The authors 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). The authors are also grateful to several unknown referees for their constructive comments on an earlier version of this article, which led to a substantial improvement of our work.
Published online: 24 Jan 2012.

A Measure of Stationarity in Locally Stationary Processes With Applications to Testing

Holger DETTE, Philip PREUSS, and Mathias VETTER

In this article 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 locally 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 a smoothing bandwidth. 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. Finally, we investigate the finite sample properties of the resulting confidence intervals and tests by means of a simulation study and illustrate the methodology in two data examples. Parts of the proofs are available online as supplemental material to this article.

KEY WORDS: Goodness-of-fit tests; Integrated periodogram; L^2-distance; Nonstationary processes; Spectral density.

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 therefore locally stationary processes provide an interesting class of models with more flexibility, as they offer 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 autocovariance 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, von Sachs, and Kroisandt (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 the article by 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 the articles by 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 a time-varying autoregressive structure or stationarity) for a time-varying spectral density has found much less attention. Sergides and Paparoditis (2009) investigated semiparametric hypotheses and proposed a bootstrap test in this context, but the rare literature typically focuses on testing for second-order stationarity, which we call stationarity throughout this article for the sake of simplicity. 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 estimating empirical wavelet coefficients by 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). Usually statistical inference of the spectrum in locally stationary processes depends on local averages of the periodogram and a common feature in many of these methods is the fact that the statistical inference depends on the choice of additional regularization parameters. For example, the (Haar) multiple test of von Sachs and Neumann (2000), checking the significance of the coefficients in a wavelet expansion of the spectral density of the locally stationary process, depends on the threshold value (and method). Similarly, Paparoditis (2009, 2010) compared nonparametric estimators of the spectral density both of the stationary and the 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

Holger Dette is Professor (E-mail: *holger.dette@ruhr-uni-bochum.de*), Philip Preuss is Ph.D. Student, and Mathias Vetter is PostDo, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany. The authors 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). The authors are also grateful to several unknown referees for their constructive comments on an earlier version of this article, which led to a substantial improvement of our work.

Journal of the American Statistical Association
September 2011, Vol. 106, No. 495, Theory and Methods
DOI: 10.1198/jasa.2011.tm10811

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 the article by 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.

In contrast to the literature cited in the previous paragraph, which has its main focus on testing stationarity, the present article is devoted to deviations from stationarity in locally stationary processes. We propose an extremely simple measure for a deviation from stationarity by determining 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

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

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 $f : [-\pi,\pi] \to \mathbb{C}$, such that the hypothesis

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

is satisfied, that is, the given locally stationary process is in fact stationary. On the other hand, if the process is not stationary, D^2 (or a corresponding standardized version) is a natural 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. The result is 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. As a consequence, we obtain a simple and intuitive tool for investigating deviations from stationarity in locally stationary processes, which includes—in contrast to the available literature—the construction of asymptotic confidence intervals and tests for precise hypotheses. As any method for time–frequency analysis, the new test proposed in this article depends on a regularization parameter which constitutes a compromise between time and frequency resolution. However, to our knowledge the new method is the only available procedure for testing stationarity in locally stationary processes which avoids a further regularization *and* yields a test statistic with a simple limit distribution. It is therefore particularly attractive for practitioners and we demonstrate that it is at least competitive to (and in many cases more efficient than) the tests which have been proposed so far.

The remaining part of the article is organized as follows. In Section 2 we introduce the necessary notation and 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 provide a comparison of the new test with the procedures proposed by von Sachs and Neumann (2000) and Paparoditis (2010) and illustrate the methodology by reanalyzing several data examples, which have recently been discussed in the literature. Finally, some more technical details required in the asymptotic analysis are deferred to the Appendix and to supplemental material on the web.

2. MEASURING STATIONARITY

Locally stationary time series can be defined via a sequence of stochastic processes $\{X_{t,T}\}_{t=1,\ldots,T}$ ($T \in \mathbb{N}$), where each observation $X_{t,T}$ exhibits a linear representation of the form

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

Throughout this article, we assume that the random variables Z_t are independent and identically normally distributed, with mean zero and variance σ^2. The assumption of Gaussianity of the errors is not necessary in general, but it is imposed to simplify technical arguments (see Remark 2).

Since the constants $\psi_{t,T,l}$ are in general time-dependent, each process $X_{t,T}$ will typically not be stationary. Nevertheless, if one assumes that the coefficients behave like some smooth functions in a neighborhood of time t/T, the time series becomes locally stationary in the sense that observations close nearby show approximately stationary behavior. Therefore we adopt not only the usual summation condition

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

but impose additionally that there exist twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$ with

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

Furthermore, we assume that the technical conditions

$$\sum_{l=-\infty}^{\infty} \sup_{u\in[0,1]} |\psi_l(u)||l|^2 < \infty, \qquad (2.3)$$

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

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

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,

$$f(u,\lambda) = \frac{\sigma^2}{2\pi} \left| \psi(u, \exp(-i\lambda)) \right|^2, \qquad (2.6)$$

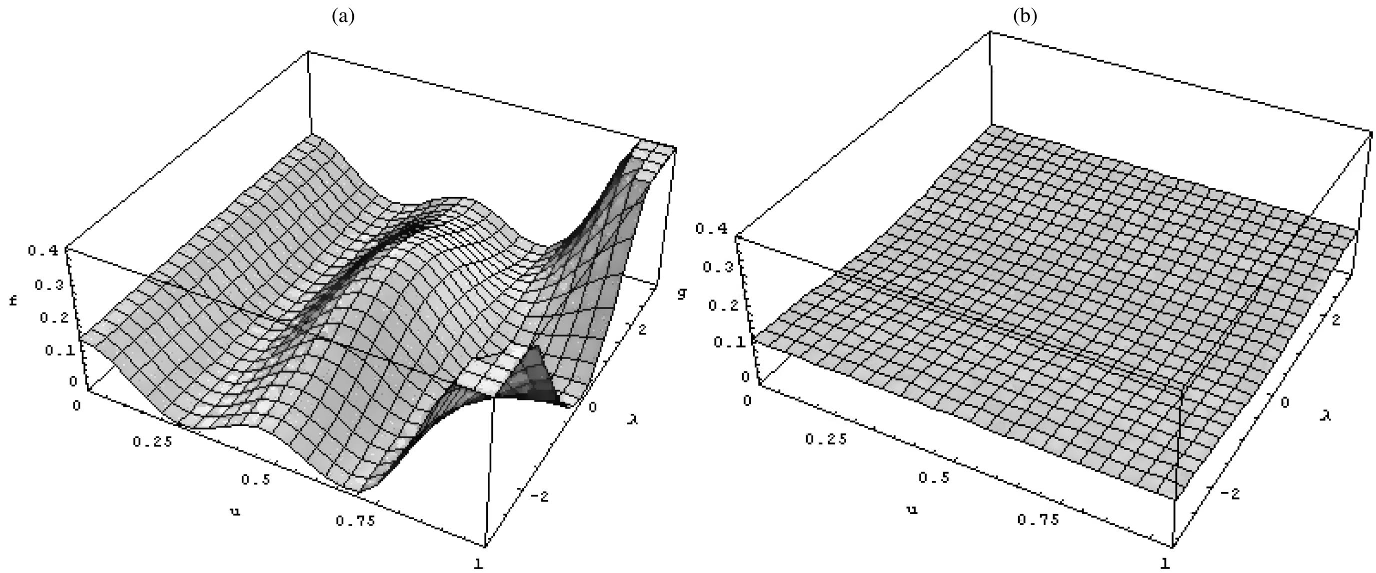

Figure 1. (a) Plot of the local spectral density $f(u,\lambda)$ in (2.7). (b) Plot of the best approximation of f by the spectral density $g(u,\lambda) = g^*(\lambda) = \int_0^1 f(u,\lambda)\,du$.

where the function ψ is given by $\psi(u, \exp(-i\lambda)) := \sum_{l=-\infty}^{\infty} \psi_l(u)\exp(-i\lambda l)$. Assumption (2.3) is in general rather mild, as it is satisfied by a variety of time-varying ARMA(p,q) models. Also, it guarantees existence of the time-varying spectral density function (2.6), and it was shown by Dahlhaus (1996) that f is unique under the assumptions stated above.

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

Lemma 1. The minimal distance defined in (1.1) is given by

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

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

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

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

Example 1. Consider the tvMA(2) process $X_{t,T} = \cos(2\pi t/T)Z_t - (t/T)^2 Z_{t-1}$, where $\sigma^2 = 1$. We obtain by a straightforward calculation

$$f(u,\lambda) = \frac{1}{2\pi}\{\cos(2\pi u)^2 - 2u^2\cos(2\pi u)\cos(\lambda) + u^4\}, \quad (2.7)$$

and the best approximation via a stationary spectral density is given by $g^*(\lambda) = \int_0^1 f(u,\lambda)\,du = \frac{7}{20\pi} - \frac{1}{2\pi^3}\cos(\lambda)$. 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 1, an estimate for it can easily be constructed by estimating the integrals

$$F_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_0^1 f^2(u,\lambda)\,du\,d\lambda, \quad (2.8)$$

$$F_2 = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\int_0^1 f(u,\lambda)\,du\right)^2 d\lambda. \quad (2.9)$$

In this context the additional regularization comes into play. 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. The main idea is to split the entire data into M blocks with N observations each, from which we define appropriate local periodograms. Precisely, let $I_N^X(u,\lambda) := |J_N^X(u,\lambda)|^2$ be the usual periodogram around time u computed from N observations, that is, we set

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

and $X_{i,T} = 0$, if $i \notin \{1,\ldots,T\}$ (see Dahlhaus 1997). 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 from appropriate Riemann approximations in time and frequency. For this purpose, we use the notation $u_j := \frac{t_j}{T} := \frac{N(j-1)+N/2}{T}$ $(j = 1,\ldots,M)$ for the mid-point of each block and set

$$\hat{F}_{1,T} = \frac{1}{T}\sum_{k=1}^{\lfloor N/2\rfloor}\sum_{j=1}^{M} I_N^X(u_j,\lambda_k)^2, \quad (2.10)$$

$$\hat{F}_{2,T} = \frac{1}{N}\sum_{k=1}^{\lfloor N/2\rfloor}\left(\frac{1}{M}\sum_{j=1}^{M} I_N^X(u_j,\lambda_k)\right)^2, \quad (2.11)$$

where $\lambda_k = 2\pi k/N$ denotes the usual Fourier frequency. The estimate of the measure of stationarity D^2 is finally given by

$$\hat{D}_T^2 = 2\pi \hat{F}_{1,T} - 4\pi \hat{F}_{2,T}. \tag{2.12}$$

Note that it is not self-evident that our approach is working, because it is usually not true that an integrated function of the periodogram converges to the corresponding integrated function of the spectral density (as it was pointed out for stationary processes by Taniguchi 1980). This phenomenon is visible here as well, but the difference only regards multiples of π, which explains the somewhat unintuitive definition of $\hat{D}_T^2$ in (2.12).

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

$$N \to \infty, \qquad M \to \infty, \qquad \frac{T^{1/2}}{N} \to 0, \qquad \frac{N}{T^{3/4}} \to 0. \tag{3.1}$$

Our first result specifies the asymptotic distribution of the vector $(\hat{F}_{1,T}, \hat{F}_{2,T})^T$ defined by (2.10) and (2.11).

Theorem 1. If the assumptions (2.3)–(2.5) and (3.1) are satisfied, then

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

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

$$\mathbf{\Sigma} = \begin{pmatrix} \frac{5}{\pi}\int_{-\pi}^{\pi}\int_0^1 f^4(u,\lambda)\,du\,d\lambda & \frac{2}{\pi}\int_{-\pi}^{\pi}(\int_0^1 f(u,\lambda)\,du\int_0^1 f^3(u,\lambda)\,du)\,d\lambda \\ \frac{2}{\pi}\int_{-\pi}^{\pi}(\int_0^1 f(u,\lambda)\,du\int_0^1 f^3(u,\lambda)\,du)\,d\lambda & \frac{1}{\pi}\int_{-\pi}^{\pi}((\int_0^1 f(u,\lambda)\,du)^2\int_0^1 f^2(u,\lambda)\,du)\,d\lambda \end{pmatrix} \tag{3.2}$$

and

$$d_{N,T} = \frac{N}{4\pi T}\int_{-\pi}^{\pi}\int_0^1 f^2(u,\lambda)\,du\,d\lambda, \tag{3.3}$$

respectively.

Proof. For a proof of the asymptotic normality in Theorem 1 we use the Cramér–Wold device and show weak convergence of the linear combination

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

for all vectors $\mathbf{c} \in \mathbb{R}^2$. This is done in several steps: First, we prove in the Appendix that the lth cumulant of the statistic $A_T(\mathbf{c})$ satisfies

$$\mathrm{cum}_l(A_T(\mathbf{c})) = O\big(T^{1-l/2}\big) \quad \text{for } l \geq 2, \tag{3.4}$$

which shows, inter alia, that the cumulants of degree higher than 2 converge to zero. Afterward, we calculate both the means and the variances and covariances of $\hat{F}_{1,T}, \hat{F}_{2,T}$ and obtain

$$\mathrm{cum}_1(A_T(\mathbf{c})) = o(1) \tag{3.5}$$

as well as

$$\lim_{T\to\infty} T\,\mathrm{Var}(\hat{F}_{1,T}) = \frac{5}{\pi}\int_{-\pi}^{\pi}\int_0^1 f^4(u,\lambda)\,du\,d\lambda, \tag{3.6}$$

$$\lim_{T\to\infty} T\,\mathrm{Var}(\hat{F}_{2,T}) = \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\left(\int_0^1 f(u,\lambda)\,du\right)^2 \times \int_0^1 f^2(u,\lambda)\,du\right)d\lambda, \tag{3.7}$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(\hat{F}_{1,T}, \hat{F}_{2,T}) = \frac{2}{\pi}\int_{-\pi}^{\pi}\left(\int_0^1 f(u,\lambda)\,du \times \int_0^1 f^3(u,\lambda)\,du\right)d\lambda. \tag{3.8}$$

Details of these computations are available online as supplemental material. The assertion then follows because the cumulants of the random variable $A_T(\mathbf{c})$ converge to the cumulants of a normal distribution with mean 0 and variance $\mathbf{c}^T\mathbf{\Sigma}\mathbf{c}$.

Now a straightforward application of the Delta-method yields the asymptotic distribution of the statistic $\hat{D}_T^2$ defined in (2.12).

Theorem 2. If the assumptions of Theorem 1 are satisfied, then we have

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

where the constant $d_{N,T}$ is defined in (3.3) and the asymptotic variance is given by

$$\begin{aligned}\tau^2 &= 20\pi\int_{-\pi}^{\pi}\int_0^1 f^4(u,\lambda)\,du\,d\lambda \\ &\quad - 32\pi\int_{-\pi}^{\pi}\left(\int_0^1 f(u,\lambda)\,du\int_0^1 f^3(u,\lambda)\,du\right)d\lambda \\ &\quad + 16\pi\int_{-\pi}^{\pi}\left(\left(\int_0^1 f(u,\lambda)\,du\right)^2\int_0^1 f^2(u,\lambda)\,du\right)d\lambda.\end{aligned} \tag{3.9}$$

The asymptotic bias $4\pi d_{N,T} = \frac{2\pi N}{T}F_1$ in Theorem 2 is nonvanishing, since the condition $N = o(\sqrt{T})$ is excluded by the assumptions in (3.1). Note that the proof of Theorem 1 shows that such a bias is inevitable for all possible growth conditions on M and N, as in general either $\hat{F}_{1,T}$ or $\hat{F}_{2,T}$ is biased. Nevertheless, it can easily be estimated by the statistic

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

It follows from Theorem 1 that

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

thus Theorem 2 yields

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

For statistical applications it remains to estimate the asymptotic variance τ^2. In general (i.e., if $D^2 > 0$), suitable estimators for the three integrals in (3.9) can be constructed from rescaled versions of

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

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

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

as the following theorem shows. Its proof is similar to the one of Theorem 1 and is therefore omitted.

Theorem 3. If the assumptions of Theorem 1 are satisfied, we have

$$\hat{\tau}_1^2 \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} \int_0^1 f^4(u, \lambda)\, du\, d\lambda,$$

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

$$\hat{\tau}_3^2 \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} \left(\left(\int_0^1 f(u, \lambda)\, du \right)^2 \int_0^1 f^2(u, \lambda)\, du \right) d\lambda.$$

By means of the preceding result, we obtain a consistent estimator for the asymptotic variance τ^2 by setting

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

Under the assumption of stationarity (i.e., $D^2 = 0$) the asymptotic variance in (3.9) reduces to

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

and one estimates $\tau_{H_0}^2$ consistently by $\hat{\tau}_{H_0}^2 = 4\pi^2 \hat{\tau}_1^2$.

Remark 1. (a) If D^2 is used as a measure for the deviation from stationarity of a locally stationary process, we obtain from Theorem 2 a consistent estimate. In order to address for different size of the data we propose to use a normalized measure of stationarity defined by

$$R := \frac{D^2}{2\pi F_1} = 1 - 2\frac{F_2}{F_1}, \tag{3.10}$$

where F_1 and F_2 are defined in (2.8) and (2.9), respectively. Note that $R \in [0, 1]$, because by the Cauchy–Schwarz inequality $2F_2 \le F_1$. In the following we use R as a measure for the relative deviation from stationarity. A natural estimator for R is given by

$$\hat{R} = \frac{\hat{D}_T^2 + B_T}{2\pi \hat{F}_{1,T}},$$

and we obtain from Theorem 1 and the Delta-method that the central limit theorem

$$\sqrt{T}(\hat{R} - R) \xrightarrow{D} \mathcal{N}(0, \rho^2) \tag{3.11}$$

holds, where

$$\rho^2 = 4\frac{F_2^2}{F_1^4}\sigma_{11} - 8\frac{F_2}{F_1^3}\sigma_{12} + 4\frac{1}{F_1^2}\sigma_{22}$$

and the σ_{ij} are the elements of the covariance matrix $\mathbf{\Sigma}$ defined in (3.2). In the same way as before we obtain a consistent estimator $\hat{\rho}_{H_1}$ for ρ, and it follows that the interval

$$\left[0, \hat{R} + \frac{\hat{\rho}_{H_1}}{\sqrt{T}} u_{1-\alpha} \right] \tag{3.12}$$

is an asymptotic $(1 - \alpha)$ confidence interval for the "parameter" R, where $u_{1-\alpha}$ denotes the $(1 - \alpha)$ quantile of the standard normal distribution. The coverage probability of (3.12) 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 hypothesis (1.2) is equivalent to $D^2 = 0$, this can be accomplished by rejecting the null hypothesis whenever

$$\hat{D}_T^2 + B_T \ge \frac{\hat{\tau}_{H_0}}{\sqrt{T}} u_{1-\alpha}, \tag{3.13}$$

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 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 $\sqrt{T}(\hat{D}_T^2 + B_T)$ 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*

$$H_0 : R > \varepsilon \quad \text{versus} \quad H_1 : R \le \varepsilon, \tag{3.14}$$

where $R \in [0, 1]$ is defined in (3.10) (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). The parameter $\varepsilon > 0$ in (3.14) denotes a prespecified constant for which the statistician agrees to analyze the data under the additional assumption of stationarity. From the weak convergence in (3.11) an asymptotic level α test for the hypothesis (3.14) is defined by rejecting the null hypothesis, whenever

$$\hat{R} - \varepsilon < \frac{\hat{\rho}_{H_1}}{\sqrt{T}} u_\alpha. \tag{3.15}$$

Note that this procedure allows for accepting the null hypothesis of "approximate stationarity" at controlled Type I error.

Remark 2. It should be noted that the results in Theorem 2 can be extended to the case where the innovations are not necessarily normally 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 2 remains valid with a different asymptotic variance, that is,

$$\tau_g^2 = \tau^2 + \frac{\kappa_4}{\kappa_2^2}\left\{4\int_0^1 \left(\int_{-\pi}^{\pi} f^2(u,\lambda)\,d\lambda\right)^2 du \right.$$
$$+ 4\int_0^1 \left(\int_{\pi}^{\pi} f(u,\lambda)\left(\int_0^1 f(v,\lambda)\,dv\right) d\lambda\right)^2 du$$
$$- 8\int_0^1 \left(\int_{-\pi}^{\pi} f(u,\lambda)^2\,d\lambda \right.$$
$$\left.\left. \times \int_{-\pi}^{\pi} f(u,\lambda)\left(\int_0^1 f(v,\lambda)\,dv\right) d\lambda\right) du\right\},$$

where τ^2 is defined in (3.9) and κ_2 and κ_4 denote the variance and the fourth cumulant of the innovations. Even though τ_g^2 is in general different from τ^2, both quantities coincide at least in the stationary case.

Remark 3. Following Dahlhaus (1997) it is too restrictive to use the more natural definition

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

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, e.g., Paparoditis 2009, 2010 or Dahlhaus 2009).

4. FINITE SAMPLE PROPERTIES

In this section we study the finite sample properties of the statistical applications mentioned in Remark 1. In particular we investigate the approximation of the level of the test for stationarity and present a detailed power comparison with two alternative methods which have recently been proposed for this purpose. We also study the coverage probability of the asymptotic confidence intervals for the parameter D^2 (or R) and the problem of validating stationarity in the sense of (3.14), and illustrate potential applications in two data examples. All reported simulation results are based on 1000 simulation runs.

4.1 Testing for Stationarity

4.1.1 Some Comments on the Choice of M and N. It is an intrinsic feature that any statistical inference in locally stationary processes depends on the choice of M and N (or on similar parameters) in the definition of the local periodogram (see Dahlhaus 1997; Paparoditis 2010; among many others) and in this paragraph we investigate the impact of this choice on level and power of the test for stationarity. For this purpose we exemplarily consider the tvMA(2) model

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

with independent, standard Gaussian distributed innovations Z_t and different choices for the parameter b. Other scenarios showed similar results and are not depicted for the sake of brevity. In Table 1 we display the simulated rejection probabilities of the test (3.13) for different combinations of M and N. The choice $b = 0$ corresponds to the null hypothesis of stationarity, while the values $b = 0.5$ and $b = 1$ represent two alternatives. We observe that all rejection probabilities are increasing with M (under the null hypothesis and alternative). In other words, a larger choice of M will usually increase the power but at the same time yields an overestimation of the nominal level. From Table 1 it is visible that in all cases the choice $M = 8$ leads to a reasonable approximation of the nominal level and only for sample sizes $T = 1024$ and $T = 2048$ a larger value can be recommended. This corresponds to intuition because 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. To obtain both

Table 1. Rejection probabilities of the test (3.13) in 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
256	16	16	0.134	0.251	0.177	0.33	0.411	0.574
512	64	8	0.053	0.125	0.117	0.234	0.430	0.582
512	32	16	0.125	0.240	0.216	0.354	0.535	0.697
1024	128	8	0.038	0.122	0.140	0.263	0.690	0.825
1024	64	16	0.079	0.166	0.224	0.381	0.758	0.866
2048	256	8	0.049	0.112	0.246	0.386	0.921	0.97
2048	128	16	0.073	0.161	0.297	0.452	0.943	0.983
2048	64	32	0.158	0.267	0.438	0.558	0.965	0.98

Table 2. Simulated level of the test (3.13) in models (4.2)–(4.4) for different parameters

			$H_0: X_t + \phi X_{t-1} = Z_t$									
			$\phi = -0.5$		$\phi = -0.25$		$\phi = 0$		$\phi = 0.25$		$\phi = 0.5$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	32	8	0.009	0.041	0.025	0.077	0.046	0.092	0.062	0.141	0.077	0.152
512	64	8	0.014	0.048	0.032	0.085	0.048	0.11	0.056	0.127	0.055	0.146
1024	128	8	0.021	0.057	0.036	0.081	0.039	0.097	0.047	0.118	0.06	0.132
			$H_0: X_t = Z_t + \theta Z_{t-1}$									
			$\theta = -0.5$		$\theta = -0.25$		$\theta = 0.1$		$\theta = 0.25$		$\theta = 0.5$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	32	8	0.054	0.135	0.058	0.132	0.041	0.086	0.02	0.08	0.014	0.048
512	64	8	0.041	0.121	0.042	0.107	0.031	0.085	0.021	0.075	0.014	0.055
1024	128	8	0.048	0.109	0.055	0.117	0.042	0.098	0.034	0.08	0.019	0.071
			$H_0: X_t + \phi X_{t-1} = Z_t + \theta Z_{t-1}$									
		$(\phi, \theta) =$	$(-0.25, -0.25)$		$(-0.5, -0.5)$		$(0.5, 0.5)$		$(0.25, -0.25)$		$(0.25, 0.25)$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	32	8	0.048	0.119	0.051	0.119	0.041	0.109	0.056	0.137	0.045	0.113
512	64	8	0.049	0.102	0.054	0.123	0.046	0.097	0.063	0.151	0.042	0.105
1024	128	8	0.047	0.105	0.056	0.12	0.05	0.106	0.053	0.124	0.052	0.093

a reasonable approximation of the level and a good power behavior of the test, we therefore recommend to choose $M = 8$ for sample sizes up to $T = 2048$. In the following section we will illustrate that this choice yields a good approximation of the nominal level for many other stationary processes.

4.1.2 Approximation of the Nominal Level. We now study the approximation of the nominal level of the test for stationarity defined in (3.13) in the following ARMA(p, q) models:

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

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

$$X_t + \phi X_{t-1} = Z_t + \theta Z_{t-1}, \tag{4.4}$$

where we investigate various combinations of the parameters ϕ and θ. On the basis of the results of the previous paragraph we used $M = 8$ in the three examples under consideration. The corresponding rejection probabilities are depicted in Table 2 for various parameters in the models (4.2)–(4.4). In most situations the approximation of the nominal level of the test is rather accurate. In the AR(1) and MA(1) model the test is conservative for the parameters $\phi = -0.5$ and $\theta = 0.5$, respectively, whereas in the case of an ARMA(1, 1) process we observe a reasonable approximation of the Type I error for all parameter combinations.

Remark 4. Note that the statistic $\hat{D}_T^2$ is essentially a quadratic form in the random variables $I_N^X(u_j, \lambda_k)$. Therefore it is reasonable to investigate an approximation of its distribution by a scaled and shifted χ^2 distribution. To be precise, let Y_T denote a χ^2 distributed random variable with $n_T = 2T/\hat{\tau}_{H_0}^2$ degrees of freedom and set $a_T = -B_T - 1$; then $\frac{1}{n_T} Y_T + a_T$ is approximately normally distributed with expectation $-B_T$ and variance $\frac{\tau_{H_0}^2}{T}$. Because under the null hypothesis the statistic $\hat{D}_T^2$ is also approximately normally distributed with the same expectation and variance, an alternative asymptotic level α test can be defined by rejecting the null hypothesis, whenever

$$\hat{D}_T^2 > \frac{\chi^2_{n_T, 1-\alpha}}{n_T} + a_T, \tag{4.5}$$

where $\chi^2_{n_T, 1-\alpha}$ denotes the $(1 - \alpha)$ quantile of χ^2 distribution with n_T degrees of freedom. In order to check if this test yields a better approximation of the nominal level we have exemplarily investigated model (4.4). The rejection probabilities of the test (4.5) are depicted in Table 3, and comparing these with the

Table 3. Simulated level of the test (4.5) in model (4.4) for different parameters

			$H_0: X_t + \phi X_{t-1} = Z_t + \theta Z_{t-1}$									
		$(\phi, \theta) =$	$(-0.25, -0.25)$		$(-0.5, -0.5)$		$(0.5, 0.5)$		$(0.25, -0.25)$		$(0.25, 0.25)$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	32	8	0.045	0.119	0.045	0.118	0.04	0.106	0.053	0.135	0.045	0.112
512	64	8	0.048	0.102	0.052	0.122	0.045	0.097	0.057	0.149	0.04	0.102
1024	128	8	0.047	0.105	0.055	0.119	0.05	0.106	0.051	0.124	0.051	0.098

Table 4. Rejection probabilities of the test (3.13) for various alternatives of local stationarity

	(4.6)		(4.7)		(4.8)		(4.9)		(4.10)		(4.11)	
T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	0.828	0.919	0.681	0.83	0.899	0.964	0.207	0.336	0.740	0.864	0.627	0.775
512	0.986	0.997	0.943	0.978	0.997	1	0.277	0.422	0.951	0.978	0.865	0.924
1024	1	1	0.999	1	1	1	0.386	0.516	0.990	1	0.980	0.997

third part of Table 2 we do not observe substantial differences with respect to the approximation of the nominal level. Because other scenarios (which we have not depicted here for the sake of brevity) showed a similar picture we will work with quantiles from the normal distribution throughout this section.

4.1.3 Power Consideration. In this section we study the power of the test and compare it with some alternative procedures for testing stationarity from the literature. For this purpose we consider the following nonstationary processes:

$$X_{t,T} = 1.1\cos(1.5 - \cos(4\pi t/T))Z_{t-1} + Z_t, \tag{4.6}$$

$$X_{t,T} = 0.6\sin(4\pi t/T)X_{t-1} + Z_t, \tag{4.7}$$

$$X_{t,T} = (0.5X_{t-1} + Z_t)I_{\{[1,T/4]\cup[3T/4+1,T]\}}(t) + (-0.5X_{t-1} + Z_t)I_{[T/4+1,3T/4]}(t), \tag{4.8}$$

$$X_{t,T} = (-0.5X_{t-1} + Z_t)I_{\{[1,T/2]\cup[T/2+T/64+1,T]\}} + 4Z_t I_{[T/2+1,T/2+T/64]}(t), \tag{4.9}$$

where $I_A(t)$ denotes the indicator function of a set A. Additionally we investigate a wavelet process (see Van Bellegem and von Sachs 2008) defined by

$$X_{t,T} = \sum_{k=0}^{T-1} w_1(k/T)\psi_{1,k-t}Z_{1,k} + \sum_{k=0}^{T-1} w_2(k/T)\psi_{2,k-t}Z_{2,k}, \tag{4.10}$$

where the functions w_1, w_2 and the coefficients $\psi_{1,k}, \psi_{2,k}$ are given by

$$w_1(x) = 0.6\cos(4\pi x), \qquad w_2(x) = 0.8x^2,$$
$$\psi_{1,k} = 2^{-1/2}I_{\{0\}}(k) - 2^{-1/2}I_{\{1\}}(k),$$
$$\psi_{2,k} = \tfrac{1}{2}I_{\{0,1\}}(k) - \tfrac{1}{2}I_{\{2,3\}}(k),$$

respectively. Our final example is of the form

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t)A(t/T,\lambda)\,d\xi(\lambda), \tag{4.11}$$

where ξ is an orthogonal-increment process and

$$A(u,\lambda) = \frac{1}{2\pi}\big(1 - \big(0.3 + 8.7\cos(6\pi u) \cdot I_{[1/2,3/4]\times[\pi/2,2\pi/3]}(u,\lambda)\big)\exp(-i\lambda)\big).$$

One can show that $X_{t,T}$ from (4.11) defines a locally stationary process with spectral density $f(u,\lambda) = |A(u,\lambda)|^2$ (see Dahlhaus 1996).

The models (4.6)–(4.8) have been considered by Paparoditis (2010) as well, and we investigate these for the sake of a comparison. Models (4.6) and (4.7) have smooth changes in the autocovariance structure while model (4.8) represents a process with a structural break. In model (4.9) the autocovariance changes only in a very short time period, while model (4.10) represents a locally stationary wavelet process (see Van Bellegem and von Sachs 2008). Finally, model (4.11) corresponds to a locally stationary process, where nonstationarity is expressed over a short period of time and a narrow frequency band. According to the discussion in Section 4.1.1 we choose $M = 8$, because this choice yields a satisfying approximation of the nominal level. The simulated rejection probabilities are displayed in Table 4. We observe that all deviations from stationarity are detected with reasonable probabilities. For the alternatives (4.6)–(4.8) the power of the new test can now be compared with the power of the test proposed by Paparoditis (2010) for checking stationarity. This author considered the statistic

$$\max_{N/(2T)\le s\le T-N/(2T)} N\sqrt{b}(Q_T(u_s) - \mu_T),$$

where μ_T is a bias which depends only on the sample size and some kernel function K, and

$$Q_T(u_s) := \int_{-\pi}^{\pi}\left(\frac{1}{N}\sum_j K_b(\lambda-\lambda_j)\left(\frac{I_N(u_s,\lambda_j)}{\hat f(\lambda_j)} - 1\right)\right)^2 d\lambda$$

is an estimate of the integrated distance $\int_{-\pi}^{\pi}(f(u_s,\lambda)/f(\lambda) - 1)^2\,d\lambda$. Here $\hat f(\lambda) = \frac{1}{T}\sum_j K_h(\lambda-\lambda_j)I_T(\lambda_j)$ denotes a kernel estimate of the spectral density (from all data) under the assumption of stationarity, $K_h(x) = \frac{K(x/h)}{h}$, and b and h are bandwidths. In other words, this approach divides the data in segments with $N = o(T)$ points, and for each segment the L^2-distance between the spectral density estimator using only N data points and the global spectral density estimator using all T data points is calculated. Finally, the maximum of the distances over all segments is determined. Paparoditis (2010) pointed out that the power of this test depends sensitively on the choice of the bandwidths in the kernel estimators and investigated several bandwidths and window lengths. The range of the power of his test for these different choices is displayed in Table 5 for the sake of an easy comparison. We observe that in models (4.6) and (4.8) the test proposed in this article yields a substantially larger power than the procedure proposed by Paparoditis (2010) for all choices of the bandwidth and window length in this procedure. Even for the best choice (which is of course not known in applications) the power of the new test is larger. In model (4.7) it is at least possible to choose the parameters of regularization such that the test of Paparoditis (2010) yields a similar power behavior as the test (3.13). However, for most choices of these regularization parameters the power of the test proposed by (Paparoditis 2010) is also substantially smaller in this example.

Table 5. The range of the power of the test of Paparoditis (2010) for different bandwidths and window lengths

	(4.6)		(4.7)		(4.8)	
T	5%	10%	5%	10%	5%	10%
256	[0.495, 0.805]	[0.675, 0.898]	[0.485, 0.75]	[0.641, 0.85]	[0.46, 0.605]	[0.555, 0.765]

An alternative test for stationarity in locally stationary processes was proposed by von Sachs and Neumann (2000). Their approach is based on Haar wavelets dividing the data and the interval $[-\pi, \pi]$ into different parts and comparing the spectral density estimators on neighboring intervals, both in time and frequency. Formally, the test rejects the null hypothesis of stationarity if the maximum of the coefficients

$$\frac{2^{(j+j')/2}}{\pi^{1/2}} \int_{k'2^{-j'}\pi}^{(k'+1)2^{-j'}\pi} \big(I_{\langle \lfloor k2^{-j}T \rfloor, \lfloor (k+1/2)2^{-j}T \rfloor \rangle}(\lambda) \\ - I_{\langle \lfloor (k+1/2)2^{-j}T \rfloor, \lfloor (k+1)2^{-j}T \rfloor \rangle}(\lambda)\big)\, d\lambda,$$

with $k = 0, \ldots, 2^j - 1$, $k' = 0, \ldots, 2^{j'} - 1$, $j = 0, \ldots, J, j' = 0, \ldots, J'$, exceeds a suitable quantile of the normal distribution, where

$$I_{\langle L,R \rangle}(\lambda) := \frac{1}{2\pi(R - L + 1)} \left| \sum_{t=L}^{R} X_t \exp(-i\lambda t) \right|^2$$

is the ordinary periodogram for the sequence $X_L, \ldots, X_R$. In Table 6 we show the simulated power of this test for the six alternatives (4.6)–(4.11). It was pointed out by von Sachs and Neumann (2000) that the performance of the test depends sensitively on the choice of the regularization parameters J and J' and in the simulation study these are chosen such that the nominal level of the wavelet test is approximated with reasonable accuracy, that is, $J = 2$ and $J' = 0$. The results in Tables 4 and 6 are therefore directly comparable. We observe that in model (4.6)–(4.8) the test (3.13) proposed in the present article yields substantially larger rejection probabilities than the test of von Sachs and Neumann (2000). Interestingly the power of the latter test does not increase with the sample size in these examples, and it is not obvious that their test is consistent in this setting. On the other hand, the wavelet test has some advantages compared to the test (3.13) for the alternatives (4.9)–(4.11) by its construction because these examples represent noncontinuous changes in the local spectral density. While the improvement is substantial in example (4.9), it is not so visible for the alternatives (4.10) and (4.11) if the sample size is larger than $T = 512$.

4.2 Confidence Intervals

The coverage probability of the confidence intervals defined in (3.12) is exemplarily investigated for the tvMA(2) model defined in (4.1) and for sample sizes $T \leq 2048$. The results are displayed in Table 7 for various values of T and M. We observe reasonable coverage probabilities for $b = 0$ and $b = 0.5$; in the case of stationarity ($b = 0$), the actual coverage probability is in fact slightly larger than the prespecified level, while the opposite behavior is observed in the case $b = 1$. In all cases the accuracy is improved with increasing sample size. Note for $b = 0.5$ and $b = 1$ that the coverage probability is increasing with M while it is relatively stable in the case $b = 0$, as expected for stationary time series. Based on our numerical experiments, 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. In particular, for sample sizes $T \leq 2048$ 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 $b = 1$). Table 7 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 again 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.

4.3 Validating Stationarity

In this section we investigate the finite sample properties of the test (3.15) for the precise hypothesis (3.14), where the bound for accepting stationarity is chosen as $\varepsilon = 0.1$. As in Section 4.2 we consider the tvMA(2) model defined in (4.1). Note that for the values $b = 1$, $b = 0.5$, and $b = 0$ we obtain $R_{b=1} \approx 0.134$, $R_{b=0.5} \approx 0.042$, and $R_{b=0} = 0$, respectively. Therefore the cases $b = 0$ and $b = 0.5$ correspond to the alternative $H_1 : R < 0.1$, while the choice $b = 1$ gives a scenario for the null hypothesis $H_0 : R \geq 0.1$. The results are depicted in Table 8 for various choices of M and N. The power of the test (3.14) is decreasing with M. We recommend to choose $M = 8$ for sample sizes $T \leq 1024$ and $M = 16$ if $T = 2048$ in order to obtain a satisfying approximation of the nominal level and a good size of the test. Note that the choice $b = 1$ does not correspond to the

Table 6. Rejection frequencies of the wavelet test for stationarity proposed by von Sachs and Neumann (2000) under various alternatives

	(4.6)		(4.7)		(4.8)		(4.9)		(4.10)		(4.11)	
T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
256	0.224	0.306	0.275	0.375	0.219	0.312	0.616	0.686	0.934	0.96	0.938	0.972
512	0.207	0.289	0.258	0.342	0.228	0.317	0.754	0.80	0.99	1	0.999	1
1024	0.198	0.294	0.20	0.294	0.232	0.344	0.94	0.964	1	1	1	1

Table 7. Coverage probabilities of the asymptotic confidence intervals (3.12) in the tvMA(2) model (4.1) for different values of b

			$b=0$		$b=0.5$		$b=1$	
T	N	M	95%	90%	95%	90%	95%	90%
256	32	8	0.975	0.932	0.948	0.9	0.876	0.819
256	16	16	0.975	0.939	0.954	0.923	0.897	0.854
512	64	8	0.966	0.924	0.954	0.904	0.885	0.815
512	32	16	0.955	0.93	0.94	0.906	0.896	0.85
1024	128	8	0.958	0.9	0.933	0.886	0.875	0.8
1024	64	16	0.959	0.924	0.954	0.907	0.888	0.827
2048	256	8	0.948	0.894	0.926	0.867	0.847	0.774
2048	128	16	0.954	0.91	0.928	0.875	0.908	0.867
2048	64	32	0.975	0.945	0.946	0.909	0.943	0.913

boundary of the hypothesis. Therefore it is expected that the level of the test for $b=1$ should be smaller than the nominal level and approximately α for a parameter b^* with $R_{b^*}=0.1$, that is, $b^* \approx 0.815$. This can be observed for sample sizes larger than 512.

4.4 Data Examples

In this subsection we illustrate the application of the developed methodology by reanalyzing two data examples from the recent literature. We begin with an example from neuroscience which has been considered by von Sachs and Neumann (2000) and Paparoditis (2009) as well. These authors analyzed a dataset of tremor data recorded in the Cognitive Neuroscience Laboratory of the University of Québec at Montreal. On the whole, there are 3072 observations, and the first-order differences $\Delta_t = X_t - X_{t-1}$ of the time series are analyzed. 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

$$\hat{f}(u,\lambda) = \frac{2\pi}{N}\sum_{k=1}^{N}\frac{1}{b}K\left(\frac{\lambda-\lambda_k}{b}\right)I_N^X(u,\lambda_k) \qquad (4.12)$$

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 nonstationarity in the data and it might be of interest to investigate this visual conclusion by the statistical methodology developed in this article. 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 the null hypothesis with a p-value of 0.033. Note that these findings confirm the investigations of 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), again the first-order differences $\Delta_t = X_t - X_{t-1}$ 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 the estimate $\sqrt{1200}\frac{\hat{D}^2}{\hat{\tau}_{H_1}} \approx 0.454$ for the standardized distance $\sqrt{T}\frac{D^2}{\tau}$. A plot of the density estimate (4.12) is shown in the right panel of Figure 2, where we used $N=134$ 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 from the result obtained by Paparoditis (2010). An possible ex-

Table 8. Rejection frequencies of the test for the precise hypothesis (3.14) (with $\varepsilon = 0.1$) of the tvMA(2) model (4.1) for different values of b

			$b=0$ (H_1)		$b=0.5$ (H_1)		$b=0.815$ (H_0)		$b=1$ (H_0)	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	32	8	0.224	0.334	0.153	0.254	0.09	0.154	0.055	0.107
256	16	16	0.178	0.254	0.131	0.2	0.071	0.113	0.051	0.084
512	64	8	0.338	0.474	0.209	0.317	0.091	0.155	0.043	0.081
512	32	16	0.316	0.433	0.201	0.292	0.084	0.139	0.04	0.068
1024	128	8	0.563	0.677	0.304	0.422	0.085	0.149	0.029	0.069
1024	64	16	0.467	0.587	0.288	0.375	0.087	0.141	0.03	0.052
2048	256	8	0.793	0.863	0.444	0.567	0.132	0.201	0.03	0.061
2048	128	16	0.715	0.797	0.435	0.547	0.07	0.13	0.008	0.026
2048	64	32	0.621	0.694	0.332	0.428	0.05	0.091	0.007	0.013

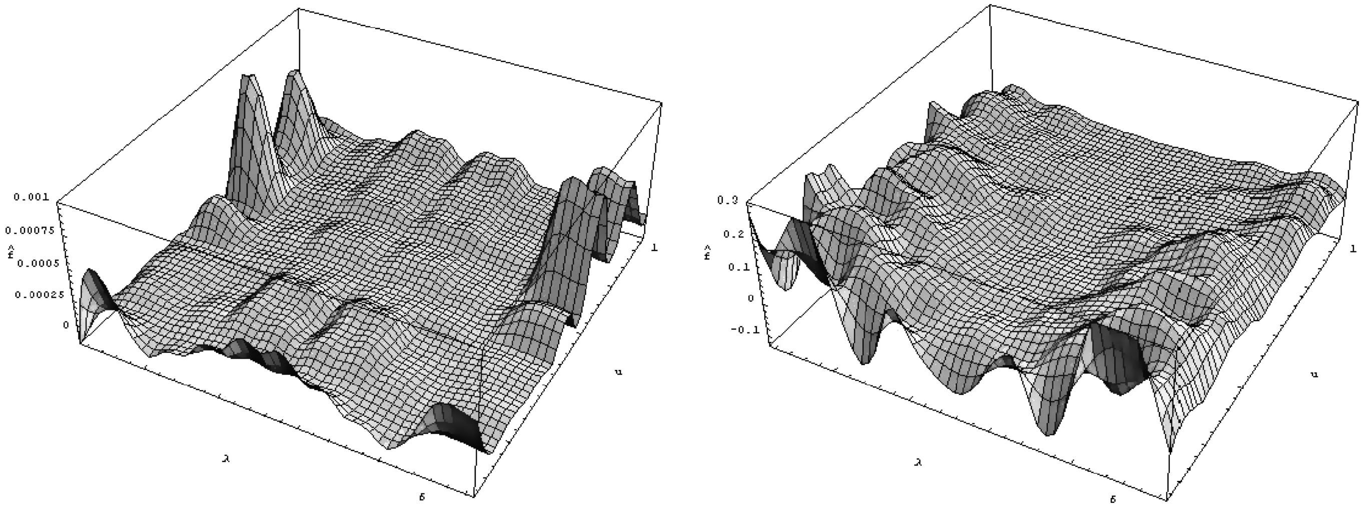

Figure 2. Estimate of the spectral density for two datasets. *Left*: neuroscience data. *Right*: egg data.

planation is that we smooth the differences between the local spectral density and the best stationary approximation over time while Paparoditis (2010) took the maximum as the test statistic.

APPENDIX: TECHNICAL DETAILS

This appendix is devoted to the proof of (3.4), which is based on the product theorem for cumulants from the work of Brillinger (1981), to which we refer for all unexplained terminology. We restrict ourselves to the case $\mathbf{c} = (1, 0)$, as the general one follows from exactly the same lines with an additional amount of notation. By assumption we focus on $l \geq 2$ only. In this case the cumulant is invariant under translation, and therefore it suffices to compute the lth cumulant of $\sqrt{T}\hat{F}_{1,T}$, which by multilinearity becomes

$$\begin{aligned}
&\mathrm{cum}_l(\sqrt{T}\hat{F}_{1,T}) \\
&\quad = \mathrm{cum}_l\left(\frac{1}{T^{1/2}}\sum_{j=1}^{M}\sum_{k=-\lfloor (N-1)/2\rfloor}^{\lfloor N/2\rfloor} I_N^X(u_j,\lambda_k)^2\right) \\
&\quad = \frac{1}{T^{l/2}}\sum_{j_1,\ldots,j_l=1}^{M}\sum_{k_1,\ldots,k_l=-\lfloor (N-1)/2\rfloor}^{\lfloor N/2\rfloor} \mathrm{cum}\big(I_N^X(u_{j_1},\lambda_{k_1})^2,\ldots, \\
&\qquad I_N^X(u_{j_l},\lambda_{k_l})^2\big).
\end{aligned}$$

Assumption (2.2) yields

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

where we have set $t_{j,p} = t_j - N/2 + 1 + p$ and $u_{j,p} = t_{j,p}/T$. It is obvious that we can forget about the small order term above.

Let us introduce some further notation. First we set

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

for $i \in \{1,\ldots,l\}$. Then we define

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

for any indecomposable partition $\nu = \nu_1 \cup \cdots \cup \nu_{2l}$ with subsets containing two elements of the table

$$\begin{matrix}
(1,1) & (1,2) & (1,3) & (1,4) \\
\vdots & \vdots & \vdots & \vdots \\
(l,1) & (l,2) & (l,3) & (l,4)
\end{matrix}$$

The product theorem finally gives

$$\mathrm{cum}_l(\sqrt{T}\hat{F}_{1,T}) = \frac{1}{(2\pi)^{2l}}\sum_{\nu} V(\nu),$$

where the summation on the right side is performed with respect to all indecomposable partitions containing two elements, due to the normality of Z. As the number C_l of such 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 ith row hooks with the $(i+1)$st one (otherwise we switch the rows accordingly).

Let us also fix $v_1,\ldots,y_l$ and j_1. That the first row hooks with the second one means that a product of the form $\mathrm{cum}(Y_{11}, Y_{23})$ appears within (A.1). In order for it to be nonzero the corresponding indices of Z have to be equal, that is, there has to exist a relation of the form

$$\begin{aligned}
&t_{j_1} - N/2 + 1 + p_1 - v_1 = t_{j_2} - N/2 + 1 + r_2 - x_2 \\
&\qquad \Leftrightarrow \qquad r_2 = p_1 - v_1 + x_2 + t_{j_1} - t_{j_2}. \qquad \text{(A.2)}
\end{aligned}$$

Thus r_2 has to satisfy both

$$x_2 - v_1 + t_{j_1} - t_{j_2} \leq r_2 \leq x_2 - v_1 + t_{j_1} - t_{j_2} + N - 1$$

and

$$0 \leq r_2 \leq N-1,$$

and since v_1, x_2, and t_{j_1} 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 nonzero cumulant. By induction it follows that given j_1 there is only a finite number D_l of valid choices for the indices $j_2, \ldots, j_l$, and in the following we keep one of these fixed as well.

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

$$p_1 - r_2 = v_1 - x_2 + t_{j_2} - t_{j_1} \tag{A.3}$$

that have to be satisfied in order for the cumulants to be nonzero. Since ν is a partition, each variable $p_1, \ldots, s_l$ appears exactly once within these $2l$ expressions. On the other hand, using

$$\sum_{k=-\lfloor (N-1)/2 \rfloor}^{\lfloor N/2 \rfloor} \exp(-i\lambda_k t) = \sum_{k=-\lfloor (N-1)/2 \rfloor}^{\lfloor N/2 \rfloor} \exp(-i\lambda_t k) = N I_{\{N\mathbb{Z}\}}(k) \tag{A.4}$$

for any integer t, we see that further l equations

$$p_i - q_i + r_i - s_i = m_i N \quad \text{with } m_i \in \mathbb{Z} \tag{A.5}$$

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 $m_1, \ldots, m_l$. 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 $p_1, q_1, r_1, s_1, \ldots, p_l, q_l, r_l, s_l$ in that particular order. Then we set

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

and

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

where the vectors v_i and w_i relate to the homogeneous versions of the equations in (A.5) and (A.3) in an obvious way: v_i refers to the conditions involving $p_i, \ldots, s_i$, whereas w_i represents the $2l$ equations from (A.3) in arbitrary order. The claim on the dimension of the solution space can be deduced from the following lemma.

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

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

$$\alpha_2 v_2 + \alpha_3 v_3 + \cdots + \alpha_l v_l + \beta_1 w_1 + \cdots + w_{2l} \beta_{2l} = 0. \tag{A.6}$$

Focus on those w_{i_1} with a nonzero 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 (A.6) would not be satisfied. Now the partition is chosen in such a way that the first row of the table hooks with the second one; thus there is a vector $w_{i_{12}}$ with one nonzero entry within rows 1 to 4 and the second nonzero 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 jth row hooks with the $(j+1)$st.

With the aid of these results the proof of assertion (3.4) is now easy. From the previous discussion we know that the sum in (A.1) has the following upper bound:

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

SUPPLEMENTARY MATERIALS

Mean and covariance: The additional material contains parts of the proof of Theorem 1, namely the computation of the asymptotic mean and (co)variance. (supplement.pdf)

[Received December 2010. Revised March 2011.]

REFERENCES

Adak, S. (1998), "Time-Dependent Spectral Analysis of Nonstationary Time Series," *Journal of the American Statistical Association*, 93, 1488–1501. [1113]

Berger, J. O., and Delampady, M. (1987), "Testing Precise Hypotheses," *Statistical Science*, 2, 317–335. [1117]

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, New York: McGraw Hill. [1123]

Dahlhaus, R. (1996), "On the Kullback–Leibler Information Divergence of Locally Stationary Processes," *Stochastic Process and Their Applications*, 62, 139–168. [1115,1120]

——— (1997), "Fitting Time Series Models to Nonstationary Processes," *The Annals of Statistics*, 25 (1), 1–37. [1113,1115,1118]

——— (2000), "A Likelihood Approximation for Locally Stationary Processes," *The Annals of Statistics*, 28 (6), 1762–1794. [1113]

——— (2009), "Local Inference for Locally Stationary Time Series Based on the Empirical Spectral Measure," *Journal of Econometrics*, 151, 101–112. [1113,1114,1118]

Dahlhaus, R., and Polonik, W. (2006), "Nonparametric Quasi Maximum Likelihood Estimation for Gaussian Locally Stationary Processes," *The Annals of Statistics*, 34 (6), 2790–2824. [1113]

——— (2009), "Empirical Spectral Processes for Locally Stationary Time Series," *Bernoulli*, 15, 1–39. [1113]

Lee, S., Ha, J., Na, O., and Na, S. (2003), "The Cusum Test for Parameter Change in Time Series Models," *Scandinavian Journal of Statistics*, 30 (4), 781–796. [1113]

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. [1113]

Neumann, M. H., and von Sachs, R. (1997), "Wavelet Thresholding in Anisotropic Function Classes and Applications to Adaptive Estimation of Evolutionary Spectra," *The Annals of Statistics*, 25, 38–76. [1113]

Paparoditis, E. (2009), "Testing Temporal Constancy of the Spectral Structure of a Time Series," *Bernoulli*, 15, 1190–1221. [1113,1118,1122]

——— (2010), "Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms," *Journal of the American Statistical Association*, 105 (490), 839–851. [1113,1114,1118,1120-1123]

Priestley, M. B. (1965), "Evolutionary Spectra and Non-Stationary Processes," *Journal of the Royal Statistical Society, Ser. B*, 62, 204–237. [1113]

Sakiyama, K., and Taniguchi, M. (2003), "Testing Composite Hypotheses for Locally Stationary Processes," *Journal of Time Series Analysis*, 24 (4), 483–504. [1113]

——— (2004), "Discriminant Analysis for Locally Stationary Processes," *Journal of Multivariate Analysis*, 90, 282–300. [1113]

Sergides, M., and Paparoditis, E. (2009), "Frequency Domain Tests of Semiparametric Hypotheses for Locally Stationary Processes," *Scandinavian Journal of Statistics*, 36, 800–821. [1113]

Taniguchi, M. (1980), "On Estimation of the Integrals of Certain Functions of Spectral Density," *Journal of Applied Probability*, 17, 73–83. [1116]

Van Bellegem, S., and von Sachs, R. (2008), "Locally Adaptive Estimation of Evolutionary Wavelet Spectra," *The Annals of Statistics*, 36, 1879–1924. [1120]

von Sachs, R., and Neumann, M. H. (2000), "A Wavelet-Based Test for Stationarity," *Journal of Time Series Analysis*, 21, 597–613. [1113,1114,1121,1122]